现代数学译丛 22

重分形：理论及应用

〔美〕戴维·哈特 著
华南理工大学分形课题组 译



科学出版社

现代数学译丛 22

重分形：理论及应用

〔美〕 戴维·哈特 著

华南理工大学分形课题组 译

科学出版社

北 京

图字：01-2012-1137 号

内 容 简 介

重分形分析是 20 世纪 80 年代以来分形几何最重要的成果，已成为分形几何的核心课题之一，它广泛应用于动力系统、湍流、降雨量模型、地震和昆虫数量的空间分布、金融时间序列模型及交通网络模型. 本书侧重将重分形分析理论应用于统计，特别是用统计学的观点来估计分形维数是其他书所未涉及的独到的贡献.

本书第一部分介绍背景和重分形测度的不同定义，特别是用格覆盖和点中心球覆盖的两种构造. 第二部分介绍大偏差下的重分形公式，主要讨论通过大偏差理论得到上述两种构造的“重分形机制”. 第三部分讨论 Rényi 维数的估计、性质及其应用. 独特的是将偏差分为内在与外在两类形式，并通过理论及实例指出：内在偏差由概率分布的内在性质引起，外在偏差由取样与所采用的统计方法形成，从而给出了一些实用的方法与技巧. 同时给出丰富的应用实例，特别详细讨论了地震位置空间点模型. 附录部分概括介绍了各种维数的定义和大偏差理论.

这是一本将重分形理论应用于统计的非常好的参考书. 可供数学及相关专业高年级本科生、研究生及科研教学人员参考.

图书在版编目（CIP）数据

重分形：理论及应用/(美)哈特(Harte, D.)著；华南理工大学分形课题组译.
—北京：科学出版社，2012
(现代数学译丛；22)
ISBN 978-7-03-034204-1
Ⅰ. ①重… Ⅱ. ①哈… ②华… Ⅲ. ①分形学-研究 Ⅳ. ①O415.5
中国版本图书馆 CIP 数据核字(2012) 第 083761 号

责任编辑：李 欣 赵彦超／责任校对：宋玲玲
责任印制：钱玉芬／封面设计：陈 敬

科 学 出 版 社出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com
骏 杰 印 刷 厂 印刷
科学出版社发行 各地新华书店经销
*
2012 年 6 月第 一 版 开本：B5(720×1000)
2012 年 6 月第一次印刷 印张：14 3/4
字数：276 000

定价：65.00 元
(如有印装质量问题，我社负责调换)

中 文 版 序

测度的重分形分析是 20 世纪 80 年代以来分形几何最重要的成果, 已成为分形几何的核心课题之一, 它研究测度的分形结构以及与其他学科, 特别是与统计物理的联系. 重分形测度及维谱的研究最早源于物理学家们对自然现象中的重分形模型的研究, 例如, 在湍流模型中它们被用来刻画高速旋涡区域的间歇性质; 在动力系统中, 它们用来描述粒子经过某吸引区域的频率; 有兴趣的读者可以参见 Mandelbrot, Frisch-Parisi, Halsey 等的相关文献. 另外由于它有非常强的应用背景, 自 80 年代引入以来, 在理论与应用上均取得非常重要的成果, 近年来发展非常迅速, 涉及的领域也越来越多.

为了解重分形分析的基本思想和方法, 先考察测度的一些最基本的信息以及这些信息间的联系. 设 μ 是平面上一个正测度, 有限正测度也称为一个质量分布, 大家首先关心的是 μ 的分布, 粗略地说, 就是 μ 在每个球上的质量, 可以从下面两个方面观测测度的分布.

***L^q* 谱, 热力学极限** 与 μ 的质量分布密切相关的是 μ 的配分函数:

$$S_\delta(q) = \sum_B \mu(B)^q, \quad q \in \mathbb{R},$$

其中 B 是长度为 δ 的边平行于坐标轴的正方形, 它们构成平面的一个分划, 上式右端对分划中所有的正方形求和. 由配分函数定义的热力学极限为

$$\tau(q) = \lim_{\delta \to 0} \frac{\log S_\delta(q)}{\log \delta}.$$

若上述极限存在, 则配分函数满足标度律, 从而知道它的增长性态. 函数 $\tau(q)$ 可以反映测度的整体性质, 通过它进而了解 μ 的分布. 此外热力学极限对 q 是否可微涉及系统是否出现相变. 但如何确定上述热力学极限 (甚至判别它是否存在) 一般说来是一个非常困难的问题.

τ 的下述变形有时在讨论中更为方便, 设 μ 是 $\mathbb{R}^n$ 上的一个给定的具有紧支撑的 Borel 概率测度, 设 $q \in \mathbb{R}$, 记

$$\tau(q) = \tau(\mu, q) = \liminf_{\delta \downarrow 0} \frac{\log\left(\sup \sum_i \mu(B_\delta(x_i))^q\right)}{\log \delta},$$

其中上确界取遍所有互不相交的半径为δ, 且中心位于μ的支撑上的闭球族$\{B_\delta(x_i)\}$. $\tau(q)$ 称为 μ 的 L^q 谱, 它是一个单调上升的凹函数. 当 $q > 1$ 时, $\tau(q)/(q-1)$ 又称为 μ 的 Rényi 维数.

维谱　测度的另一个重要信息是测度的局部结构, 即测度的密度. μ 在点 x 的 α- 维密度定义为

$$\lim_{r\to 0}\frac{\mu(B(x,r))}{(2r)^{\alpha}},$$

其中 $B(x,r)$ 表示以点 x 为中心, r 为半径的球. 在此情形, 对于很小的 r, $\mu(B(x,r)) \approx r^{\alpha}$, 因此, 若知道 μ 在每一点的密度, 那么 μ 的分布就清楚了. 我们知道, 均匀分布在每点的密度相同, 例如具有正测度的 Lebesgue 测度是均匀的, 它在几乎所有的点的密度为 1. 当一个分布不均匀时, 一个自然的想法是将密度为 α 的点, 作成一个集合 $E(\alpha)$. $E(\alpha)$ 的大小用它的维数 $\dim E(\alpha)$ 来度量 (例如对于一片树林, $E(\alpha)$ 为高度为 α 的树的集合, 尽管我们不知道这些树长在什么位置, 但我们知道这些树有多少). 通过这种方式, 可以将一个复杂的测度的支集分解为支撑均匀分布的集族. 测度的维谱 (或称奇异谱) 定义为 $f(\alpha) = \dim E_{\alpha}$, 其中 $\dim_H E_{\alpha}$ 表示集 E_{α} 的 Hausdorff 维数.

确定测度在一点的密度是一个非常困难的问题, 下面介绍一种目前在重分形分析中经常遇到同时也是更为容易处理的情形. 设 n 为一个正整数, μ 是 $\mathbb{R}^n$ 上的一个具有紧支撑的 Borel 概率测度. 设 $x \in \mathbb{R}^n$, 记

$$d(\mu,x) = \lim_{r\to 0}\frac{\log \mu(B_r(x))}{\log r},$$

如果上述极限存在, 则称 $d(\mu,x)$ 为测度 μ 在点 x 处的点态维数 (或局部维数). 如果不存在, 分别用 $\overline{d}(\mu,x)$, $\underline{d}(\mu,x)$ 表示相应的上、下极限, 并称它们为 μ 在点 x 处的上、下点态维数. 点态维数的大小反映了测度的局部分布性态. 任意点的点态密度相同的测度称为单测度, 例如均匀分布的测度和自相似测度. 点态维数不为常数的测度称为重分形测度, 为了进一步分析这类测度的点态维数的分布规律, 如同前面讨论密度的情形, 对于 $\alpha \geqslant 0$, 定义

$$E_{\mu}(\alpha) = \{x \in \mathbb{R}^n : d(\mu,x) = \alpha\}, \quad f_{\mu}(\alpha) = \dim_H(E_{\mu}(\alpha)),$$

称 $f_{\mu}(\alpha)$ 为 μ 的 Hausdorff 维谱. 这样, 整个空间 $\mathbb{R}^n$ 被分解为 $\mathbb{R}^n = \bigcup_{\alpha\geqslant 0} E_{\mu}(\alpha)$. 由于 μ 在 $E_{\mu}(\alpha)$ 上的任一点的局部分布性态相近, 我们希望知道 E_{α} 的“大小”、变化规律以及它们与 μ 整体性质的联系. 这正是测度的重分形分析的主要研究内容之一.

基于与上面同样的思想, 可以考虑测度与某一参数 α 相联系的集合 $E_{\mu}(\alpha)$, 称为 μ 对于参数 α 的水平集, 此外也可以用填充维数, 计盒维数代替 Hausdorff 维数, 考虑相应的填充维谱与计盒维谱, 从而使得重分形分析的处理对象非常广泛. 例如 $E_f(\alpha)$ 可以是函数 f 的 Hölder 指数为 α 的点的集合; 又如对于数据处理中的灰度图, $E_{\mu}(\alpha)$ 可以是灰度为 α 的子图.

Legendre 变换与重分形机理 前面提到对于某些测度, 满足两个幂率, 即 $\mu(B(x,r)) \propto r^{\alpha}$ 和 $S_{\delta}(q) \propto \delta^{\tau(q)}$. 这两个幂率之间是否存在联系, 更进一步, 在维谱与热力学极限之间是否存在联系. 物理学家 Halsey 等在 1986 年发现对于某些测度, 维谱正好是热力学极限的 Legendre 变换:

$$f(\alpha) = \inf_{-\infty<q<\infty}(\tau(q) + \alpha q),$$

而且上述变换的逆变换也成立, 亦即建立了热力学极限与维谱的联系, 同时建立了统计物理与分形几何间的联系, 作为一柄锐利的双刃剑, 重分形分析的强大威力也于此得到体现. 若测度 μ 满足的维谱满足上述 Legendre 变换, 则称 μ 满足重分形机理. 什么样的测度满足重分形机理是重分形分析的一个最基本的问题, 现在除了知道自相似测度、拟 Bernoili 测度、Gibbs 测度等熟知的测度满足重分形机理, 还获得了相当一般的结果, 但如何有效判断一个测度是否满足重分形原理还是一个没有完全解决的深刻的数学问题. 另一方面, 重分形分析近年不但在物理中的许多问题, 如湍流、相变理论、图像识别等研究中已获得重要的成功, 而且在工程、化学、生物诸学科也有重要的应用.

目前已经有许多关于分形几何在各个学科的应用的著作, 该书侧重将重分形分析理论应用于统计, 特别是用统计学的观点来估计分形维数有独到的贡献. 另一个值得重视的内容是, 作者针对统计学家, 将偏差分为内在与外在两类形式, 并分析了这两类形式对维数的影响, 他通过一些理论的分析与一些具体的实例指出内在偏差由概率分布的内在性质引起, 外在偏差则是由取样与所采用的统计方法形成, 从而给出了一些实用的方法与技巧, 着重介绍了 Rényi 维数的估计以及维数的应用. 作为著名的统计学家与应用概率学专家, 作者给出丰富的应用实例, 特别在第 9 章详细讨论了地震位置空间点模型. 这是一本将分形几何, 特别是将重分形分析理论应用于统计的非常好的参考书.

该书由华南理工大学数学系分形几何及动力系统课题组成员共同翻译, 他们中有的长期从事分形几何与重分形的理论研究, 有的为该领域很活跃的青年学者及博士, 译者的学术背景不但保证了译文的质量, 而且增色不少.

文志英

2012 年 1 月

前　言

重分形理论本质上根源于概率论, 尽管也有很多思想来自于物理、数学、统计. 它广泛应用于动力系统、湍流、降雨量模型、地震和昆虫数量的空间分布、金融时间序列模型及交通网络模型.

由于起初对地震产生的空间点模型分形维数的计算比较感兴趣, 作为统计学家和应用概率学家, 我曾经接触过这门学科. 由于重分形这门学科吸收了许多学科的理论, 并被应用于很多不同的领域, 所以不可避免地就会在专业术语、概念、技术水平等方面产生一些困难. 我试图把这些领域的思想融合在一起, 放在概率和统计的背景中, 使用统计学家能够理解并对其有用的语言. 特别地, 我的目的是为 Rényi 分形维数估计的统计性质的讨论提供一个框架.

不能说这本书只是针对统计学家的. 事实上, 从统计学的观点来估计分形维数是其他书所未涉及的. 我尝试着把偏差的形式分为两类: 内在的和外在的, 并描述了它们对维数估计的影响. 内在偏差是由概率分布的内在性质引起的, 而外在偏差是指由取样和其他方法性的困难所形成的特征, 将通过已知的数学和统计模型给出这些偏差的例子.

这本书的重点是重分形测度. 与单尺度的自相似随机过程相比, 多尺度随机过程 (有时称为“重分形”随机过程) 近期的发展是本书内容的次要方面, 这些“重分形”随机过程仅作简单的介绍.

本书第一部分给出介绍性的内容和重分形测度的不同定义, 特别地, 是通过格点覆盖和点中心球覆盖的构造给出的. 第二部分将证明这两种构造的所谓“重分形机制”可通过标准的概率技术, 即大偏差理论来得到. 最后一部分是 Rényi 维数的估计 (包含 2 阶及更高阶) 及其性质的讨论, 也讨论维数估计的各种应用, 详细说明地震位置空间点模型. 附录概括介绍维数的一些定义和大偏差理论的结果.

人们不可能希望一本书可以囊括统计学所需要的所有信息, 事实上这也不是我的目的. 附录部分包括了集合维数的各种定义及大偏差理论的一些结果. 不管怎样, 如果需要进一步的信息, 也有大量书籍和文献可供查阅. 书中我也会特别指出一些. Falconer (1990) 从几何的角度对分形做了精彩的综述, 他阐述了各种维数的定义及它们之间的关系. Falconer (1997) 的书中有更多新的东西和各种用来研究分形的数学理论方面的技巧, 本书侧重于测度方面的技巧. Ruelle (1989) 给了动力系统方面的非常好的介绍, 这些是重分形理论研究的动力来源. Ellis (1985) 对大偏差理论进行了详细的论述. Abarbanel (1955) 从物理学的角度处理了非线性的数据分析, 是

一本非常值得去读的书.

我很大程度上采用了下列书及杂志的一些定义、定理叙述及其包含的一些内容, 出于版权方面的需要, 将本书涉及上述内容的详细章节列在这里, 分别是: Mandelbrot (1989), *PAGEOPH*, §1.8; Cutler (1991), 《统计物理杂志》, §2.4.3, §2.7.5, §2.7.7; Holley & Waymire (1992), 《应用概率年刊》, §6.3.2, §6.3.3, §6.3.4, §6.3.5, §6.3.11, §6.3.12, §6.3.14; Cutler (1997), 《场论研究所通讯》, §10.5.1(2, 3), §10.5.2, §10.5.3, §10.5.4; Falconer (1990), Chichester, Wiley, 附录 A; Eills (1984), 《概率年刊》, §B.3.8, §B.3.14, §B.3.17; Eills (1985), New York, Spring-Verlag, §B.3.1, §B.3.4, §B.3.5, §B.3.6, §B.3.11, §B.3.15. 我对以上作者及本书中所涉及的作者表示感谢, 下列图像经允许后进行了修改, 引自于 Harte (1998),《非线性科学杂志》所涉及的章节分别为: §1.6, §1.8, §1.7, §9.3, §11.4, §11.2, §11.1 及 §11.3.

本书于 1993 年开始于 Wellington 的 Victoria 大学的研读团队, 团队的成员有 David Vere-jones 教授、Robert Davies 博士、Thomas Mikosch 博士和 Qiang Wang 博士. 我们的兴趣在于分形维数的估计和解释, 特别是在地震和气象学的应用领域. 在这里我要感谢团队的所有成员的帮助、为解释各种维数图形所付出的努力及自 1993 年以来所有成员的持续的兴趣, 同时也感谢 Peter Thomson 的鼓励及对这个计划感兴趣, 特别感谢 David Vere-jones 这么多年的帮助和鼓励.

David Harte

2001 年 5 月

目　　录

第四部分　附　　录

符 号 表

符号	含义
§	章节符号
$\ltimes$	连接, 例如, $(\omega_1, \cdots, \omega_n) \ltimes t = (\omega_1, \cdots, \omega_n, t)$
$\cdot\vert\cdot$	子词, 例如, $\omega\vert n$ 是序列 ω 的前 n 个字母
#	基, 例如, $\#A$ 表示集合 A 中的元素个数
$\varnothing$	空集
$\Vert\Vert$	$\mathcal{L}^\infty$ 或最大模 (除非特别说明)
$\lfloor\rfloor$	向下取整, 例如, $\lfloor t \rfloor$ 是不超过 $t \in \mathbb{R}$ 的最大整数
$\lceil\rceil$	向上取整, 例如, $\lceil\rceil$ 上不小于 $t \in \mathbb{R}$ 的最小整数
$\overset{d}{=}$	分布意义下相等, 例如, $W \overset{d}{=} U$ 表示 W 和 U 有同样的分布, $\{X(t)\} \overset{d}{=} \{Z(t)\}$ 表示 $X(t)$ 和 $Z(t)$ 具有相同的有限维分布
$\xrightarrow{\exp}$	指数收敛
$\mathbf{1}()$	特征函数, 例如, $\mathbf{1}(A) = 1$ 若条件 A 成立, 否则, 其值为零
$\mathcal{B}(\chi)$	χ 的 Borel 集
$B_{\delta_n}(k)$	格点覆盖中大小为 δ_n 的第 k 个盒子
$C(q)$	重新度量的累积生成函数
$\mathcal{D}()$	函数的定义域, 例如, $\mathcal{D}(C) = \{q : C(q) < \infty\}$
$\tilde{D}_q$	基于格点的 Rényi 维数
D_q	基于点中心的 Rényi 维数
D_q^*	级联过程的 Rényi 维数
$\dim_B$	计盒维数
$\dim_H$	Hausdorff 维数
$\dim_P$	Packing 维数
$\tilde{F}(y)$	基于格点构造的划分
$F(y)$	基于点中心构造的划分
$F^*(y)$	Moran 级联过程的划分
$F_Y(y)$	q 阶相互距离的概率分布函数
$\tilde{f}(y)$	基于格点的重分形谱
$f(y)$	基于点中心的重分形谱
$f^*(y)$	级联过程的重分形谱
$\mathcal{H}^s(F)$	集合 F 的 s-维 Hausdorff 测度

$I(y)$	大偏差理论中的熵函数
iff	当且仅当
i.i.d.	独立同分布
int	集合的内部
$J_n(\omega)$	χ 的非重叠子区间, $\omega \in \Omega^n$
$\mathbb{N}$	自然数集或计数
$o()$	收敛阶比 $\cdots$ 更高的
$O()$	收敛阶至少与 $\cdots$ 一样
$\tilde{\theta}(q)$	基于格点的关联指数
$\theta(q)$	基于点中心的关联指数
$\theta^*(q)$	级联关联指数
$\phi(y)$	$F_Y(y)$ 的非幂律分量
$\mathcal{P}^s(F)$	集合 F 的 s-维 Packing 测度
$\mathbb{R}$	实数集
$S_\delta(x)$	中心为 x 半径为 δ 的闭球
supp	测度的支撑, 也即是说, supp(μ) 是余集的 μ 测度为零的最小闭集
χ	$\mathbb{R}^d$ 的 Borel 子集
χ_δ	$\{x \in \chi : \mu[S_\delta(x)] > 0\}$
$\mathbb{Z}$	整数集

插图列表

第一部分

引言和预备知识

第 1 章　动机和背景

1.1　引　　言

本章将对本书以后章节涉及到的许多主要概念作一个直观介绍. 在 §1.2 中, 将给出分形集和重分形测度的差别, 分形和重分形方法已经被广泛地用来描述动力系统 (将在 §1.3 介绍), 具有重分形特征的过程的一个显著特征是各种相关概率分布有幂定律. 另一些应用领域是湍流、降雨量和地震模型, 这些将在 §1.4, §1.5 和 §1.6 中相应地作介绍, 这些领域中的幂定律特征在以前的文献中已经被广泛地讨论过.

本章有意与后面的其他章节不同, 重点是作描述性的和一些常用术语的介绍, 特别地, 维数和分形这两个概念将被粗略地使用, 许多概念的正式定义将在接下来的章节中给出. 集合的各种维数及它们之间的一些内在关系将在附录 A 中给出, 细节请参考 Falconer (1990).

1.2　分形集和重分形测度

Mandelbrot (1977, 1983) 开始考虑描述具有极不正则形状的对象, 他的例子包含银河系、海岸线的长度、雪花和 Cantor 集. 这些对象中, 有些有相当奇异的特征 (至少在初始时), 例如, 无穷长的海岸线和面积无穷大的雪花. 这种情形在集合不正则的时候更容易发生, 进一步地, 要求集合具有不同层级之间除了一个比例因子外本质上是相同的特征, 这样, 不规则性在越来越精细的层级上不断地被重复, 这些集合被称为自相似的 (参看附录 §A.1).

描述这些集合大小的一种方法是计算它们的分形维数, 例如, 一个不规则海岸线的维数可能大于 1, 但是小于 2, 这表明它不同于简单的直线, 且在平面上有空间充满的特征, 同样地, 雪花的表面的维数可能大于 2 但小于 3, 这表明它的表面比正则几何形状更复杂, 是通过部分体积来填充的. 关于什么是分形及分形维数的定义现在还没有被广泛地接受, 直观地, 分形集是极不正则, 而且有分数维数的对象. 但是, 我们对分形集不是那么感兴趣, 更感兴趣的是支撑在这样集合上的测度, 通常这些测度均为概率测度. 下面的例子更能清晰地说明这些.

地震的发生可能有类似分形的特征, 地震本质上发生在地壳断裂的断层处. 把一个三维的物体简单地切开就得到一个二维的平面, 但是, 考虑小的断层分叉出大

的断层, 从这些小的断层, 甚至更小的断层被发现来看, 这种复制被重复很多次到越来越精细的层级. 如果这个假设是对的话, 人们可以期望这个维数可以大于 2 且小于 3, 也就是说, 在一个大的断层的附近, 断裂应该有一些填充特征, 但是, 在断层所组成的网络里, 有一些区域比其他区域更活跃, 也即是发生地震的概率更大, 故可以认为可能发生地震的区域是一个分形集合, 在这个分形集上有一个描述事件可能性的概率测度, 通常这个概率分布是不规则的, 没有密度的. 一个有趣的问题是: 如何刻画和描述这样一个概率分布? 这是本书的最根本的主体之一.

有概率测度的分形集合的例子很容易构造出来, 一个最简单的例子就是 Cantor 测度.

1.2.1　例子——Cantor 测度

Cantor 集是通过从单位区间挖去中间的三分之一, 然后在保留下来的两个子区间中分别挖去相应中间的三分之一, 这样的过程无限地进行下去就构造了 Cantor 集, 也就是说,

$$\begin{aligned}
\mathcal{K}_0 &= [0,1],\\
\mathcal{K}_1 &= \left[0,\frac{1}{3}\right] \cup \left[\frac{2}{3},1\right],\\
\mathcal{K}_2 &= \left[0,\frac{1}{9}\right] \cup \left[\frac{2}{9},\frac{1}{3}\right] \cup \left[\frac{2}{3},\frac{7}{9}\right] \cup \left[\frac{8}{9},1\right]
\end{aligned}$$

等. Cantor 集即是 $\mathcal{K} = \bigcap\limits_{n=0}^{\infty} \mathcal{K}_n$. 注意到 $\mathcal{K}_n$ 包含两个 $\mathcal{K}_{n-1}$ 尺度的部分, 也即是

$$\mathcal{K}_n = \left(\frac{\mathcal{K}_{n-1}}{3}\right) \cup \left(\frac{2}{3} + \frac{\mathcal{K}_{n-1}}{3}\right), \quad n = 0,1,\cdots.$$

可以看到, Cantor 集包含了单位区间中 3 进制展式中不含有数字 1 的所有的数.

如何来描述 Cantor 集的大小呢? 可以看到 $\mathcal{K}_n$ 的 Lebesgue 测度是 $\left(\frac{2}{3}\right)^n \to 0 (n \to \infty)$. 另一种描述大小的方式就是计算它的维数, 维数 (记为 d_0) 的基本思想与覆盖住人们感兴趣的集合的覆盖数目有关, 比如, 用宽为 $\frac{1}{3}$ 的方盒来覆盖 $\mathcal{K}_1$ 所要的个数是 2, 用宽为 $\frac{1}{9}$ 的方盒来覆盖 $\mathcal{K}_2$ 所要的个数是 4, 等. 也就是说, 令 $N_\delta(\mathcal{K})$ 为用宽为 δ 的方盒来覆盖集合 $\mathcal{K}$ 所需要的盒子的个数, 则

$$\frac{\log N_{\delta_n}(\mathcal{K})}{-\log \delta_n} = \frac{\log(2^n)}{-\log 3^{-n}} = \log_3 2,$$

其中 $\delta_n = 3^{-n}$. 数字 $\log_3 2$ 是 Cantor 集的维数 (盒维数和 Hausdorff 维数), 可以看到覆盖所要求的盒子的个数与维数有一定的尺度关系, 也就是说, 当 $n \to \infty$ 时, $N_{\delta_n}(\mathcal{K}) \sim \delta_n^{-d_0}$.

上面关于维数的描述与计盒维数尽管不是完全一致, 但它们之间有非常密切的相关性 (参看定义 A.3.1), 维数有很多种定义方法, 差别在于这些覆盖是否是不相交的或重叠的, 用盒子还是用球覆盖, 用固定宽度还是宽度可变的 (不超过 δ), 及当 $\delta \to \infty$ 时取极限的方式, 但是, 计算覆盖的个数的基本思想是一样的. 附录 A 给了一个概要, 进一步的内容可参考 Falconer (1990).

现在通过在每部分的子区间上分配质量或概率来扩展前面的例子, 在这个例子中, 将把区间上存在的概率的 $\frac{2}{3}$ 分配给图 1.1 的右边的部分, 将 $\frac{1}{3}$ 分配给左边部分. 由构造可知, Cantor 集是闭的, 因此是这个测度的支撑, 故这个支撑的维数为 $\log_3 2$. 但是, 维数不依赖于如何分配概率. 因此, 如何描述这个支撑在 Cantor 集上的概率分布? 比较明确的是这样的分布都没有密度, 因为集合的 Lebesgue 测度为 0.

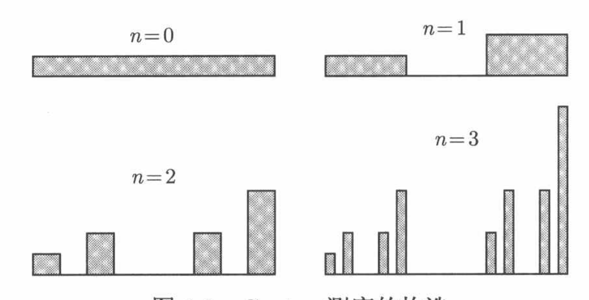


图 1.1 Cantor 测度的构造

左区间分配 $\frac{1}{3}$ 的概率和右区间分配 $\frac{2}{3}$ 的概率的 Cantor 测度的构造

考虑对 $\mathcal{K}_n(n = 1, 2, \cdots)$ 的最左边部分的概率变化率进行量化, 情形如下: 每个部分的宽度是 $\delta_n = 3^{-n}$, 第 n 步的概率测度记为 μ_n, 则对所有的 n,

$$\frac{\log \mu_n([0, 3^{-n}])}{\log \delta_n} = \frac{\log(3^{-n})}{\log(3^{-n})} = 1.$$

$\mathcal{K}_n$ 的最右边部分也可以类似地刻画, 对所有 n, 有

$$\frac{\log \mu_n([1 - 3^{-n}, 1])}{\log \delta_n} = \frac{\log((2/3)^n)}{\log(3^{-n})} = 1 - \log_3 2 \approx 0.3691.$$

现在将上述描述扩展到所有的子区间, 令 $K_n(y)$ 为宽为 δ_n 的子区间 J 的集合, 且满足若 $J \in K_n(y)$, 则

$$\frac{\log \mu_n(J)}{\log \delta_n} = y.$$

记 $\#K_n(y)$ 为在 $K_n(y)$ 中长为 δ_n 的子区间的个数, 从图 1.1 可以看出, 当 $n = 3$ 时, $\#K_3(0.3691) = 1, \#K_3(0.5794) = 3, \#K_3(0.7897) = 3$ 和 $\#K_3(1) = 1$. 这样要求

覆盖且宽为 $\delta_3 = 3^{-3}$ 的盒子数, 比如 $K_3(0.7897)$ 是 3. 现在考虑使得 $\#K_n(y) > 0$ 的那些 y 的值, 定义

$$\tilde{f}_n(y) = \frac{\log \#K_n(y)}{\log \delta_n}.$$

图 1.2 画出了当 $n = 3, 5, 10, 20$ 时的函数 $\tilde{f}_n(y)$. 函数 $\tilde{f}(y) = \lim_{n\to\infty} \tilde{f}_n(y)$ 被称为重分形谱, 该图像也在图 1.2 中画出了. 如果该测度在每个部分都进行平均地分配, 也就是说, 左右两边的概率均为 $\frac{1}{2}$, 则这个重分形谱将仅仅在一点存在, 也就是 $\tilde{f}(\log_3 2) = \log_3 2$. 这与图 1.2 中图的最大值不吻合, $\tilde{f}(y)$ 的两种形式的差别反映了这两种极限概率分布的差别. 有很多关于 $\tilde{f}(y)$ 的解释, 包括概率收敛的速率和类盒维数, 这些将在下一章进行讨论.

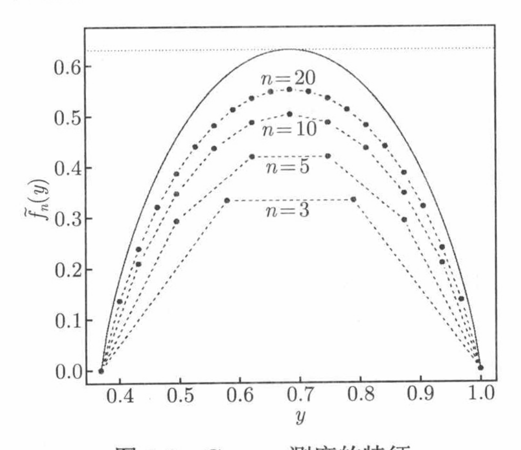


图 1.2 Cantor 测度的特征

左区间分配 $\frac{1}{3}$ 的概率和右区间分配 $\frac{2}{3}$ 的概率的 Cantor 测度的特征, 实心点对应的是 $n = 3, 5, 10, 20$ 的 $\tilde{f}_n(y)$, 实线表示的是 $\lim_{n\to\infty} \tilde{f}_n(y)$, 虚线画的是 Cantor 集的维数, 也就是 $\log_3 2$

这些概率分布也可以通过一族维数来描述, 比如 Rényi 维数, 这种维数也跟重分形谱有关. 在第二部分将讨论这些维数的理论上的性质, 在第三部分将讨论 Rényi 维数的估计方法.

不正则的集合通常对不同方法定义的维数其相应的值也不同, 比如, Hausdorff 维数、计盒维数和 Packing 维数 (参看相应的定义 A.2.3, A.3.1 和 A.4.2). 从概念的角度来说, 计盒维数非常好, 但它有一个不好的性质, 一个集合与其闭包有相同的维数. 例如, 实轴上的有理数集的盒维数为 1. 尽管 Hausdorff 维数比较难处理, 但它有更好的数学性质. 关于不同维数的定义及其之间的关系可以在附录 A 中找到.

分形集最早的定义是 Mandelbrot 在 1977 年给出的, 他定义 Hausdorff 维数比拓扑维数大的集合为分形集. 1986 年, Taylor 作进一步的讨论, 他建议修改为

Hausdorff 维数与 Packing 维数相等的集合, 如果这两者不相等, 那么这样的集合不正则以至不能被描述. 但是, 通过上面的定义, 一些不规则的集合也不是分形, 正如 Stoyan.D 和 Stoyan.H 在 1994 年指出的, 具有正面积的不规则集 (如岛) 不是分形, 尽管它的边界可能是分形 (也就是海岸线).

分形和重分形的本质差别是前者研究的是一个集合, 而后者研究的是一个测度. 像分形集一样, 一个重分形测度可能是极不正则的, 可能有很多不同阶的奇异性, 这种测度可能是也可能不是支撑在一个分形集上. 一个问题是: 是否能够把支撑集分割成一些分形集, 使得每个分割的小块上测度的奇异的阶都相同, 也就是测度是齐次的或在每个分割块上是一致分形的? 如果是可能的, 将称其为某种意义下的重分形, 这里的某种意义就是“分形”的含义, 本质上这也正是例 1.2.1 所做的.

我们能够将上面所讨论的内容统一到一个概率空间 $(\chi, B(\chi), \mu)$, 为了需要, 假设 $\chi \subset \mathbb{R}^d$, $B(\chi)$ 是 χ 上的 Borel 子集. 在例 1.2.1 中, $\chi = [0, 1)$, Cantor 测度 μ 的支撑的盒维数和 Hausdorff 维数均为 $\log_3 2$, 这个支撑的维数与概率的分配方式无关. 我们的兴趣特别集中在下述情形: 在 $B(\chi)$ 上的概率测度 (记为 μ) 是特别不正则的, 分布是不可微的, 且有很多阶不同的奇异性, 这样的测度被称为是重分形的.

重分形理论发展的动力很大程度上起源于物理文献, 重分形理论的一些应用将在以后的章节中讨论, 特别地应用在动力系统、湍流、降雨量和地震中. 在 §1.8 中, 将概括介绍它的主要特征, 这一章的最后将对本书后面部分作一个方向性的介绍.

1.3 动 力 系 统

动力系统能够表示成一个迭代映射的形式或者微分方程的集合, 但后者可以按照下面的方式重新写成一个迭代映射的形式. 令 χ 为 $\mathbb{R}^d$ 的一个 Borel 子集, 考虑变换

$$T_\xi : \chi \to \chi,$$

其中 ξ 是修改变换的参数向量. 设 $x(t_0) \in \mathbb{R}^d$ 是一个起始位置, 则 $x(t_n) = T_\xi^n(x(t_0))$ $(n = 1, 2, \cdots)$ 形成了一个确定的序列, 也就是所谓的迭代映射. 在离散时间迭代映射中, $t_n = n$, 如果时间是连续的, 可以通过下述微分方程的形式描述

$$\frac{dx}{dt} = W_\xi(x(t)),$$

其中 $t \in \mathbb{R}, x(t) \in \mathbb{R}^d$, ξ 是一个固定的参数向量, 这样的过程可以通过一个迭代映射来逼近, 令 $h = t_{n+1} - t_n$ 充分小, 则

$$x(t_{n+1}) = T_\xi(x(t_n)) \approx x(t_n) + hW_\xi(x(t_n)), \quad n \in \mathbb{Z}. \tag{1.1}$$

动力系统通常会满足下述中的一个或多个特征:

1. 若 $x(t_0)$ 有一个很小的扰动 ε, 相应的轨道

$$x'(t_n) = T_\xi^n(x(t_0) + \varepsilon)$$

可能是发散的, 而且可能与起点在 $x(t_0)$ 的轨道有很大的不同, 这个系统被称为是混沌的.

2. 许多系统可能满足这样的性质: 如果从充分长的时间看, 包含演化系统轨道路径的集合对许多不同的起始点 $x(t_0)$"看起来"是一样的. 进一步, 轨道路径仍在这个集合中, 也就是说, 这个集合中的点被映回到这个集合中, 这个集合被称为吸引子, 关于这些在 Eckmann 和 Ruelle (1985) 中有更精确的描述, 轨道路径可能是周期的, 但这个假设不是必要的.

3. 变换 T_ξ 通常有另一个不稳定性的水平, 随着 ξ 的改变, 系统的特征可能通过一列分叉点来改变, 这里吸引子的形状和其他特征也会有相应的改变.

动力系统是确定的, 至少从理论上说是这样的. 若已知 $x(t_0)$, 就能计算将来任一点的位置. 但是, 从实际观点来看, 这种计算是不可能的, 因为计算机的计算是有限的. 这是很有趣的, 因为在混沌系统 ($x(t_0)$ 作一个非常小的扰动, 则相应的轨道将有很大的差别) 中, 轨道将很不同, 但吸引子却是一样的.

如果 $x(t_0)$ 是未知的, 过程将被认为是随机的. 给定吸引域中的点 $x(t_0)$(也就是最终移动到吸引子中), 那么这个粒子一定可以在吸引子的某处找到. 我们能够描述粒子在集合 A 中的概率吗, 也即是 $\mu(A)$? 也即是说, 我们是否能观察随着时间演化的过程, 能否使用这个来描述测度 μ 的空间特征? 在这种情况下, 我们需要假设测度在变换下是不变的, 即 $\mu = T_\xi^{-1}\mu$, 系统是遍历的, 也即使得时间平均对 μ 几乎处处的 $x(t_0) \in \chi$ 均收敛.

如下面的例子所看到的, 有一部分吸引子是常常被访问的, 也有一部分不是. 事实上, 有支撑在一个集合上的测度 μ, 这个集合可能很不正则, 测度本身可能也很不正则. 描述吸引子大小和 μ 的空间特征的一种方式是计算各种"广义维数", 这是一种首先被物理学家使用和发展的方法 (Grassberger, 1983; Procaccia, 1983).

剩下的部分将对动力系统中的一些例子作简短的介绍.

1.3.1 Cantor 映射

考虑一个由 0 和 2 组成的无穷序列 (或者实验结果), 用符号记为 $\cdots, \omega_{-2}, \omega_{-1}, \omega_0, \omega_1, \omega_2, \cdots$. 每个 ω_n 是独立的, 取 0 的概率为 p_0, 取 2 的概率为 $p_2 = 1 - p_0$. 令

$$x(n) = (0.\omega_n\omega_{n+1}\omega_{n+2}\cdots)_3,$$

其中 $n \in \mathbb{Z}$, 右边表示 $x(n)$ 的基为 3 的展式 (三进制展式). 则 $x(n+1)$ 通过一个移位算子与 $x(n)$ 联系起来, 也即是说

$$x(n+1) = 3x(n) - \lfloor 3x(n) \rfloor,$$

其中 $\lfloor x \rfloor$ 是不超过 x 的最大整数. 这定义了 $[0,1]$ 上的一个离散迭代映射, 这个映射被迭代无穷多次, 例 1.2.1 中的 Cantor 测度描述了过程中访问单位区间的子集的时间的相对比率.

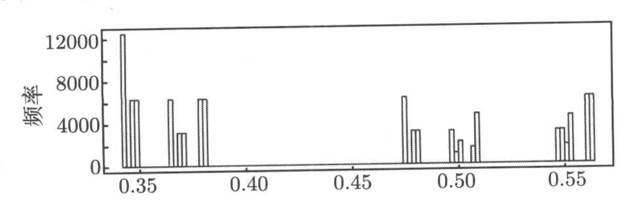
1.3.2 Logistic 映射

Logistic 映射 $T_\xi : [0,1] \to [0,1]$ 是离散的, 定义为 $T_\xi(x) = \xi x(1-x)$, 此映射给出了一个递推关系式

$$x(n+1) = \xi x(n)(1 - x(n)), \quad 0 \leqslant \xi \leqslant 4.$$

若 x 为 T_ξ 的 p-周期点, 也即是 $T_\xi^p(x) = x$ 时, p 是满足这种性质的最小整数, 则 x 被称为稳定的 (或不稳定的), 若 $|(T_\xi^p)'|$(也即是雅可比行列式) 小于 (或大于)1. 稳定点吸引附近的轨道, 不稳定点将拒绝它们. 当 ξ 增加到 $\xi_\infty \approx 3.57$ 时, 这个映射有一个有趣的分叉序列产生, 当 $0 < \xi < \xi_\infty$ 时, T_ξ 有很多不同的行为: 0 是一个不稳定的不动点、非零的稳定不动点、或者一个稳定的 2^q 周期轨道, 其中 $q \in \mathbb{Z}^+$, $\xi = \xi_q$ 满足 $\xi_q < \xi_{q+1} < \xi_\infty$. 当 $\xi = \xi_\infty$ 时, 吸引子是 Cantor 型的, 参看图 1.4. 吸引子是 T_ξ-不变集, 没有对初始条件的依赖性 (没有混沌), 其 Hausdorff 维数的估计值是 $0.532\cdots$ (Falconer, 1990, 第 173 页).

在图 1.4 中, 每次迭代出来的点在两边跳来跳去, 也就是说, 先在 0.7 到 1.0 区域的某个地方, 接下来就在 0.3 到 0.7 的某个地方等. 我们进行了 200,000 次的模拟迭代, 图 1.3(最上端柱状图) 给出了落在 0.4 到 0.6 之间的点的个数的柱状图, 由于观测图像的尺度不同, 可以看到有某些区间出现了很多点, 另外有些区间则没有点. 更具体地说, 有三个点集中的区间, 或者说是点簇, 任两个点簇之间的距离至少是 0.03. 图 1.3 的第二个图是第一个点簇的放大, 在第三个和最后一个图中, 第一个点簇进一步地被放大, 在每一个放大的图中, 注意到整体结构都是一样的, 两个不同点簇之间相应的分离及每个点簇中相应的点的个数都是相同的, 将这一点与图 1.1 中 $p_0 = \dfrac{1}{3}$ 时的 Cantor 测度的尺度特征相比, 那里每一次迭代的概率之间的差越来越大. 相似的尺度对介于 0.7 和 1.0 之间的点也同样发生, 图 1.3 中的柱形图表明不变测度支撑在一个尺度参数逼近 6 的自相似集上.



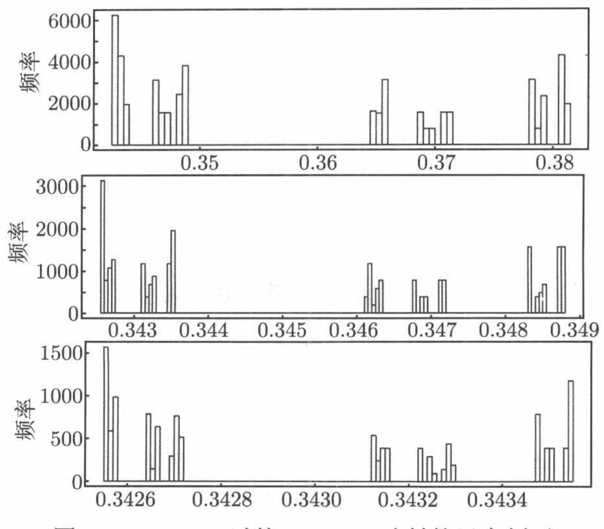


图 1.3 $\xi = \xi_\infty$ 时的 Logistic 映射的尺度刻画

$\xi = 3.569945672$ 时的 Logistic 映射的尺度特征柱状图, 第二个柱状图用更精细的比例尺画出了第一个图中点簇 < 0.4 的情形, 第三个和第四个柱状图类似地画出, 注意到尺度比例接近于 $\frac{1}{6}$

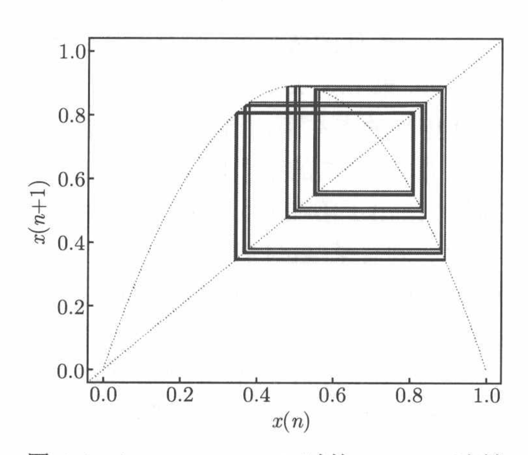


图 1.4 $\xi = 3.569945672$ 时的 Logistic 映射

$\xi = 3.569945672 \approx \xi_\infty$ 时的 Logistic 映射, 由 $x(0) = 0.8$ 出发迭代 1500 次, 仅仅画出了当 $t = 1001, \cdots, 1500$ 的 $x(t)$, 给出了充分长时间至稳定到吸引子中一个轨道的过程

May (1976, 1987) 建议使用 Logistic 映射作为一种生物人口模型, 特别地是一代一代地灭绝的模型, 这种模型适合于昆虫, 它们在冬天灭绝. Moran (1950) 在更广阔的环境中讨论了这种迭代的数学模型, 他也用生物背景解释了这种过程什么时候是稳定的.

1.3.3 Lorenz 吸引子 (Lorenz, 1963)

Lorenz 时间演化 $x(t)=(x_1(t),x_2(t),x_3(t))\in\mathbb{R}^3$ 是通过下述方程定义的

$$W_\xi(t)=\frac{dx}{dt}=\frac{d}{dt}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}\xi_1(x_2-x_1)\\\xi_3x_1-x_2-x_1x_3\\x_1x_2-\xi_2x_3\end{pmatrix}. \tag{1.2}$$

图 1.5 是通过方程 (1.1) 所给出的逼近方法画出的, 且使用一阶差分来估计导数 (Euler 方法). 取 $\xi_1=10,\xi_2=\dfrac{8}{3},\xi_3=28$, 轨道集中在一个具有高度复杂形式的吸引子上. 该吸引子由两个螺旋轨道圆盘组成, 它是混沌的, 呈现出分形, 其维数估计值接近 2.06. 在三维中旋转该图像可以看出圆盘相当的薄, 这与维数接近 2 是一致的.

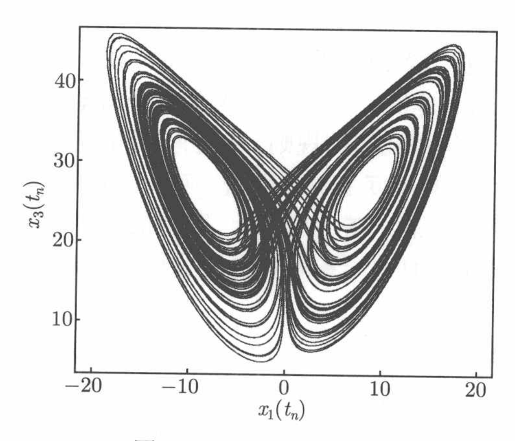


图 1.5 Lorenz 吸引子

在方程 1.1 中取 $h=0.01$, 对点 (1,1,1) 进行变换, 画出了从 1001 到 11000 次迭代的轨道路径

Lorenz (1963) 希望对流体的热对流建立模型, 对流体从下面加热, 在上边界处制冷, 然后就落下来, 这样就会在圆柱筒内传送. Abarbanel (1995) 给出了一个很好的关于该方程的背景方面的讨论及其在气象中的应用, 他注意到经常用于方程分析的参数值与对大气的行为进行建模时有效的参数值有很大的不同. Falconer (1990) 也从大气模型的观点对这些方程给了一个简单的介绍. 进一步可参考 Lorenz (1993), Eckmamm 和 Eckmann (1985, 第 622 页), Ruelle (1989, 第 9 页).

其他很多动力系统在许多文献中也被广泛讨论, 比如 Hénon 映射 (Falconer, 1990; Ruelle, 1989), 面包师的映射 (Falconer, 1990), Rössler 吸引子 (Ruelle, 1989), Ikeda 映射 (Abarbanel, 1995) 和 Kaplan-Yorke 映射.

关于动力系统和混沌方面的文献是浩大的, 有一些是非数学的描述, 一些书含有很多轨道路径和吸引子的十分漂亮的图片, 还有一些技术性更强的书. Lorenz

(1993) 对混沌系统给了一个非常好的概述, 更完善的介绍可参考 Ruelle (1989), Rasband (1990), Abarbanel (1995) 和 Falconer (1990). Eckmann 和 Ruelle (1985) 是一个重要的综述性文章, 该文给出了更深层次的细节, Cvitanović (1993), Ott 等 (1994) 和 Hao (1990) 收集了很多重要文章, Hao (1990) 编写的那一卷也包含了很多扩充的文献, Ott 等 (1994) 也含有一些介绍该主题的准备性章节. 更多的数学方面的文献可参考 David Ruelle 与其合作者一起的一些文章, 关于他的文章的文集可以在 Ruelle (1995) 中找到, Isham (1993) 从统计的角度给出了一个好的介绍, Chatterjee 和 Yilmaz (1992) 总结了动力系统在科学的很多不同分支的应用, Berliner (1992) 讨论了确定混沌系统和随机系统的关系.

1.4 湍 流

为重分形理论发展提供物理直观的一个主要领域是关于湍流中能量耗散的描述, Falconer (1990, §18.3) 提供了一个水从水龙头平滑地或慢慢地流出的例子. 随着水流量的增加, 水流变得混乱和不规则, 伴随有不同尺度的、速度变换的漩涡. 级联模型基于动能在一个大的尺度上引入系统 (例如, 暴风雪、搅拌一碗水), 但是在小尺度上只能以热的形式耗散. 在小尺度上, 黏性的作用、粒子间的摩擦力变得重要了. 这些模型都假设能量通过一系列减小的漩涡所耗散, 直到它充分小到能以热的形式耗散.

Monin 和 Yaglom (1971) 给出了关于一个很好的湍流理论发展的历史介绍, 下面很多内容都曾在该书中出现. 湍流理论起源于 19 世纪末期 Reynolds 的工作, Reynolds 数定义为 $\mathcal{R} = UL/\nu$, 其中 U 和 L 分别刻画流中的速度和长度, ν 是流体的动力黏性. 因此, $\mathcal{R}$ 是流体中作用的惯性力和粘滞力的比值, 惯性力使能量从大尺度分量转移到小尺度分量 (多相性), 而粘滞力可以磨光小尺度的多相性, 因此 $\mathcal{R}$ 值充分小的流将是薄片状的, 充分大的可能是湍流.

在 20 世纪 20 年代, Richardson 发展了一种定性的结果, 他假设发展的湍流由许多不同阶的漩涡组成 (也就是紊乱性或多相性). 由于更大漩涡稳定性的缺失, 漩涡产生了, 相应地, 也失去它们的稳定性, 生成更小的漩涡. 在这一过程中, 它们的能量被转移, 因此形成一个级联型的过程, 一旦尺度变到充分小, 则 $\mathcal{R}$ 充分小, 因此就有了很薄的流, 过程中动能转化为热能.

20 世纪 30 年代, Taylor 引入齐次和各向同性湍流的概念, 即满足下述条件: 有限空间–时间点上的流体力学量的所有有限维概率分布在任何正交变换下是不变的. 但在实数情形, 齐性和各向同性的假设不能满足 (例如边界条件), 它们对 Reynolds 数充分高的小尺度情形时的性质提供了一个有用的描述.

记 $\varepsilon(x,t)$ 为在位置 x 时间 t 处的每单位体积和单位时间内的能量消失的速度,

进一步地假设湍流已经及时地到达一个稳定态, 则在中心为 x, 半径为 δ 的球面上每单位时间消失的能量 $S_\delta(x)$ 为

$$\mu[S_\delta(x)] = \frac{3}{4\pi\delta^3}\int_{S_\delta(x)} \varepsilon(x,t)dx.$$

Kolmogorov (1941) 主张具有很高 Reynolds 数的任意湍流的充分小振幅的统计物理条件将变为齐次和各向同性的, 且一个周期时间内是稳定的. 因此他假设任意区域内每单位时间的平均能量消耗是个常数, 记这个平均值为 $\bar{\varepsilon}$. 但是, 注意到 $\varepsilon(x,t)$ 在自然中是随机的, 因此能量消耗的测度 $\mu[S_\delta(x)]$ 也是随机的.

Kolmogorov (1941) 进一步主张在 $\mathcal{R}$ 充分大时, 具有充分小尺度的速度分支的统计物理条件仅仅由 $\bar{\varepsilon}$ 和 ν 所决定, 他主张粘滞力仍然起作用的最大尺度是 $l=(\nu^3/\bar{\varepsilon})^{1/4}$. 因此, 存在一个大于 l 很多倍且远小于 L 的区间, 在这些区间中统计物理条件仅仅由单个参数 $\bar{\varepsilon}$ 决定. 记 $U_{ij}(x)$ 为 x 点 $\overrightarrow{x_ix_j}$ 方向的速度分量 (随机变量), Kolmogorov 得到对任意点 x_1, x_2,

$$E[|U_{12}(x_1)-U_{12}(x_2)|^2] = c(\bar{\varepsilon}|x_1-x_2|)^{2/3},$$

其中 $l \ll |x_1-x_2| \ll L$, c 是一致常数.

紧接着, 有人主张随着 δ 的减小, 能量消耗的变差 $\mu[S_\delta(x)]$ 无限制地增加, Kolmogorov (1962) 修改以前的"2/3 次方律"为假设 $\mu[S_\delta(x)]$ 有一个正态分布, 该分布的方差是 x 的函数且随着 δ 的减小而增加, 这就相当于把 μ 看做随机的, 这种说法将在后面关于降雨量的讨论中用到.

Monin 和 Yaglom (1971) 对湍流理论的历史发展作了详细介绍, 进一步的讨论可参考 Mandelbrot (1974), Paladin 和 Vulpiani (1987), Meneveau 和 Sreenivasan (1991), Bohr 等 (1998), Frisch (1991), Mandelbrot (1998), 及 Froedlander 和 Topper (1961) 和 Hunt 等 (1991) 中所涉及的论文集.

1.5 降 雨 量

Schertzer 和 Lovejoy (1987), Lovejoy 和 Schertzer (1985, 1990) 和 Gupta 和 Waymire (1990, 1993) 最近将与湍流相似的理论应用到降雨量领域. Gupta 和 Waymire (1990) 写道: *一般地来看, 高强度降雨量组成的簇可以嵌入到低强度小尺度区域的簇, 依次地再嵌入到依旧是低强度但大尺度区域的簇, 再嵌入到某个天气尺度最低强度降雨量的区域.*

Gupta 和 Waymire (1990) 给出在 §1.4 中概括介绍的 Kolmogorov (1962) 论述的一种推广, 他们假设对任意给定的 x, 随机测度 μ 满足下述更一般的尺度形式

$$\mu[S_{r\delta}(x)] \stackrel{d}{=} W(\delta)\mu[S_r(x)], \tag{1.3}$$

其中 $\delta \leqslant 1$, $l \ll r \ll L$, $\stackrel{d}{=}$ 表示概率分布意义下相等, $S_\delta(x)$ 是圆心在 x, 半径为 δ 的球, $W(\delta)$ 是 δ 的一个随机函数. 他们证明了 $W(\delta)$ 能够被刻画为

$$W(\delta) = \exp\{Z(-\log\delta) + \beta\log\delta\},$$

其中 $Z(t)$ 是具有平稳增量的随机过程, β 是大于 0 的任意数.

1.5.1 简单尺度

如果对所有的 t, 过程 $Z(t) = 0$, 则 $\mu[S_\delta(x)]$ 满足一个简单尺度关系

$$\mu[S_{r\delta}(x)] \stackrel{d}{=} \delta^\beta \mu[S_r(x)]. \tag{1.4}$$

令 $r = 1$, 在两边取期望 (对 μ) 得到下面等式

$$\log E[\mu^q[S_\delta(x)]] = q\beta\log\delta + \log E[\mu^q[S_1(x)]],$$

其中 $q \in \mathbb{R}$. 假设 $E[\mu^q[S_1(x)]]$ 有限, 则对充分小的 δ, 有

$$\frac{1}{q-1}\frac{\log E[\mu^{q-1}[S_\delta(x)]]}{\log\delta} \approx \beta. \tag{1.5}$$

1.5.2 例子——布朗乘子

Kolmogorov (1962) 和 Oboukhov (1962) 考虑的模型是方程 (1.3) 的一个特殊情形, 他们假设 $Z(t) = \sigma B(t)$, 其中 $\sigma > 0$, $B(t)$ 是布朗运动. 现在考虑下述启发式的论述, 假设 $W(\delta)$ 与 $\mu[S_r(x)]$ 相互独立, 令 $r = 1$. 在两边取期望 (关于 μ 和 W) 得到下面等式

$$\log E[\mu^q[S_\delta(x)]] = \log E[\exp(q\sigma B(-\log\delta) + q\beta\log\delta)] + \log E[\mu^q[S_1(x)]],$$

其中 $q \in \mathbb{R}$. 假设 $E[\mu^q[S_1(x)]]$ 有限, 则对充分小的 δ, 有

$$\frac{1}{q-1}\frac{\log E[\mu^{q-1}[S_\delta(x)]]}{\log\delta} \approx -\frac{(q-1)\sigma^2}{2} + \beta. \tag{1.6}$$

方程 (1.5) 和 (1.6) 的左边都有 q 阶 Rényi 维数的形式, 该维数的定义将在第 2 章中给出. 从方程 (1.5) 可以看出, 在简单尺度情形, 所有的 Rényi 维数都是一样的. 从方程 (1.6) 可以看出, 当尺度包含一个随机分量时, Rényi 维数都不是一样的, 相似的行为在例 1.2.1 中的 Cantor 测度情形也同样发生. 在例 2.2.1 中将证明, 如果概率是像图 1.1 那样不均匀地分配时, 则 Rényi 维数将不同, 如果概率是均匀分配的 (也就是每个子区间分配 $\frac{1}{2}$), 则 Rényi 维数都是相同的.

在这一部分的随机测度和例 1.2.1 中构造的 Cantor 测度之间有些细微的差别, Cantor 测度通过一种确定的迭代方式构造出来, 而随机测度更复杂, 是通过一种迭代的随机的方式构造出来的, Cantor 测度是多项测度族的一种特殊情形, 这将在第 3 章更全面地讨论. 在上面讨论的降雨量和湍流的例子中的随机测度是一些随机模型的例子, 这些将在第 6 章中更全面地讨论.

方程 (1.4) 给出的简单尺度关系与自相似随机过程 (Samorodnitsky 和 Taqqu, 1994) 的关系很相似, 一个随机过程 $X(t)$ 被称为是自相似的, 如果它的有限维分布满足尺度方程

$$X(\delta t) \stackrel{d}{=} \delta^H X(t) \tag{1.7}$$

对所有的 $\delta > 0, t \in \mathbb{R}, 0 < H < 1$ 成立 (进一步的细节参看 §10.6.2). 这类过程的一个例子是分式布朗运动的增量 (Mandelbrot 和 Van Ness, 1968). 当 $H > \frac{1}{2}$ 时, 过程显示出长时间的依赖性 (Beran, 1994); 当 $H = \frac{1}{2}$ 时, 分式布朗运动的增量是简单白噪声; 当 $H < \frac{1}{2}$ 时, 有一种负自相关的短时间的依赖性. 与方程 (1.3) 中尺度关系式相类似的更一般形式为

$$X(\delta t) \stackrel{d}{=} W(\delta) X(t),$$

关于满足该关系式的随机过程最近也得到了研究. 这种过程不是本书的主要内容, 只在 §10.6 中简短地讨论一下.

1.6 地 震 模 型

世界上许多比较大的城市都坐落在地震活跃带, 20 世纪, 地震造成了巨大数目的人员死亡. 比如, 比较近的 1976 年, 接近 240,000 人死于中国唐山地震中 (1976 年 7 月 28 日), 另一些造成大量人员死亡的有旧金山 (1906 年 4 月 18 日, 3,000 人死亡), 东京 (1923 年 9 月 1 日, 140,000 人死亡), 墨西哥城 (1985 年 9 月 19 日, 8,000 人死亡), 伊兹米特、土耳其 (1999 年 8 月 17 日, 17,000 人死亡).

地震预测在过去的 100 年里是一直比较活跃的研究领域, 在 20 世纪 70 年代, 有大量乐观主义的观点认为个别地震事件是可以预测的, 这与地球结构理论 (板块构造论) 和地震事件的很好记载都得到了很大的发展是相一致的. 但是这种乐观主义很快就消失了, 现在一些科学家认为这个问题非常复杂, 以至于单个地震事件是不能被预测的 (Aki, 1989; Kagan, 1997; Geller, 1997; Geller et al., 1997; Wyss et al., 1997; Kagan, 1999). 但是, 一些人仍然希望能够有效地预测相应的发生概率, 包括等高图, 就像天气图中的等高图, 在不同的区域或地区, 比较一个事件发生

的概率比给定值大的地方. Vere-Jones (1995) 给出一个地震预测的综述, Vere-Jones (2000) 从统计观点给出地震学的一个简短的介绍, 这方面更一般的介绍可参考 Lay 和 Wallace (1995) 和 Scholz (1990).

描述分裂过程的数学模型比描述天气系统的演化和行为要相对落后, 一些模型假设一直有“基本位错”发生, 大量位错的周期性发生将导致一些进一步的基本位错. 如果这些级联足够的大, 地震将通过一个充分敏感的地震网络被发现.

描述地震有很多指数律的关系, 这些都是有事实支撑的, 包括事件的震级分布 (Gutenberg-Richter 律) 和大的主震之后事件随着时间的延迟 (Omori 律). 估计由地震生成的空间点模式的分形维数的直观动机是模式可能有某种意义的自相似性, 也就是说, 簇可能在越来越细水平的簇上重复 (比如可参看图 6.1). 尽管一些簇可能比其他的要更活跃, 同样地 Cantor 测度在它的支撑上也不一定是一致的, 一般认为是大的分裂发生在大的断层处. 最引人注目的就是板块边界, 在主要断层系统中, 有一些更小的断层, 从一些更小的断层网络中偏离, 再一次以一种可能性生成网络中某类自相似层级. 在自然界中, 维数估计首先用在地震理论中, 如果地震过程确实显示出分形的特征, 则分裂过程和预测事件概率的模型也将会显示出相似的分形特征.

但是, 计算点模式的维数时有一个内在的矛盾, 理论上说有限点集的维数为 0, 因此, 维数估计描述了什么特征? 这个问题依赖于人们大脑中存在的底空间模型, 这些将在第 11 章中进一步讨论.

Wellington 地震目录包含了新西兰中部区域的一些事件, 在第 11 章的分析中, 我们选择了震级 $\geqslant 2$, 介于东经 173.6° 和 176°, 南纬 42.0° 和 40.4° 之间, 时间在 1978 年 1 月 1 日到 1995 年 12 月 31 日之间, 目录被 Wellington 地质学和核科学研究所保存 (Maunder, 1994).

地球表面由大的板块组成, 例如北美洲板块、欧亚板块、太平洋板块和澳洲板块 (Sphilhaus, 1991). 世界上大部分地震发生在板块边界的附近, 是由于一个板块相对于另一板块的运动所形成的, 新西兰位于澳洲板块和太平洋板块的边界上, 在 Wellington 地区, 太平洋板块向下拉澳洲板块, 也就是说, 太平洋板块处于澳洲板块之下, 图 1.6 中该事件的两条线粗略地画出了下降中的太平洋板块的冲突边界的位置, 可以看到深度 $\geqslant$ 40km 的大部分事件与下降过程相关, 而那些更浅的事件有一个更广泛的分布, 用 40km 作为 Wellington 目录中浅和深事件的一个边界, 在 5, 12, 33 千米处发生的事件的线画出了不是非常明确深度的浅事件.

图 1.7 和图 1.8 分别是浅和深事件的震中图, 在图 1.8 中下降过程也是显而易见的, 其中更深的事件趋向于在西部方向发生, 也注意到浅事件出现的更加集中, 在区域中更宽泛地蔓延.

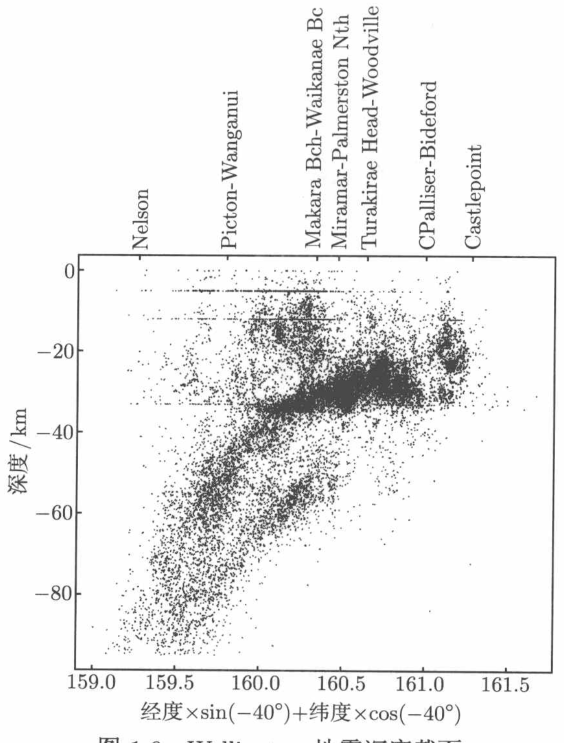


图 1.6 Wellington 地震深度截面

1985 年到 1994 年 Wellington 发生地震位置的深度截面, 其中震级 ⩾ 2, 深度 < 95km. 图中包含了 15410 个事件, 图片展示了太平洋板块下拉澳洲板块, 视角是 (接近于) 从西南往东北

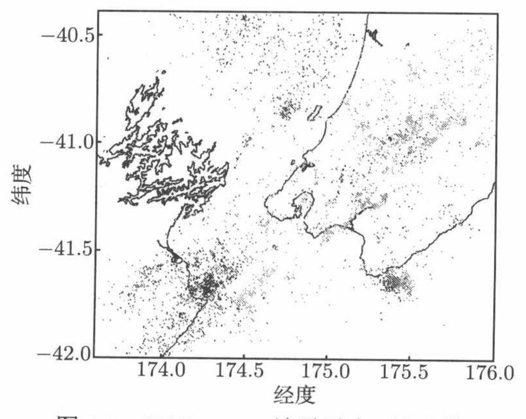


图 1.7 Wellington 地震震中：浅事件

1985 年到 1994 年之间的震级 ⩾ 2 和深度 < 40km 的 Wellington 地震震中, 地震最深的事件在灰色的最亮的阴影里, 最浅的事件是颜色最暗的, 图中包含了 10801 个事件

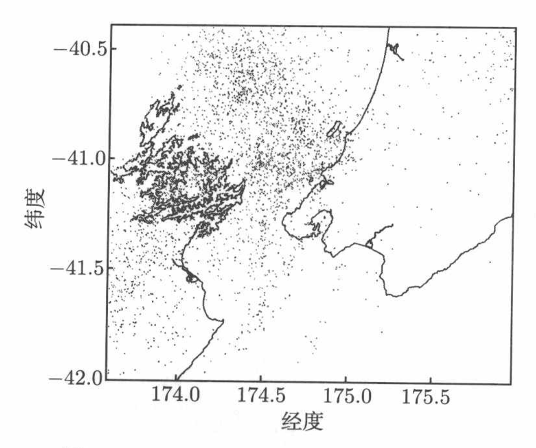


图 1.8　Wellington 地震震中：深事件

1985 年到 1994 年之间的震级 ⩾ 2 和深度 ⩾ 40km 的 Wellington 地震震中，最深的事件在灰色的最亮的阴影里，最浅的事件是最暗的，图中包含了 4952 个事件

在使用实际数据估计分形维数时有很多问题，例如，地震位置有位置上的误差，这些在分析的区域中可能不是齐次的，也有边界效应，该效应是由于在地震网络中不可能精确观察太遥远的事件所引起的，这些问题在第 11 章中将被全面介绍.

1.7　其他应用

重分形测度的许多应用是在物理过程方面，比如湍流、动力系统、降雨量和地震，Paladin 和 Vulpiani (1987) 给出了重分形在物理中许多领域的应用，其中也包含湍流，Scholz 和 Mandelbrot (1989) 包含一些分形思想在地球物理中的应用方面的论文.

一些人也假设全局气候被一个相对低维的混沌动力系统所决定 (Nicolis 和 Nicolis, 1984)，一个被观察的时间序列，像 Wellington, NZ 每天的最高温度，可以看成是这种动力系统向一个低维的观察空间作投影，使用这样一个观察数据，是否有可能决定全局系统的维数？这个问题在 §10.4.4 中将给一个简短的讨论，如果这个过程不是确定的，而是随机的，则可以期望它有无穷多个自由度 (参看 §10.5). 通过使用局部温度数据，Wang (1995) 尝试着确定是否有证据表明这些数据是由相对低维的系统所生成，他的结论说明关于降维的估计能够通过统计偏差来解释.

平行于重分形测度理论的发展，自相似随机过程 (方程 (1.7)) 也发展起来了，这个过程的增量能够有一个稳定的分布律，或者可能有长时间的依赖性，各种各样的辅助过程 (平交道等) 具有分形特征，这些将仅仅在 §10.6 中作简短介绍，这些过程已经用来建立金融数据模型 (Mandelbrot, 1997) 和网络交通模型 (Willinger et al.,

1995; Resnick, 1997; Willinger 和 Paxson, 1998; Park 和 Willinger, 2000).

自相似随机过程不是本书讨论的重点, 因为这些模型最关心的是它们的长时间依赖性和重尾特征, 与辅助过程相关的分形性质仅仅是次要方面. 进一步地说, 自相似随机过程仅仅满足一个单尺度规律. 最近, 自相似随机过程的单尺度特征已经被扩展到多尺度过程, 这种过程被称为重分形随机过程, 这些将在 §10.6.5 作简短介绍.

1.8 重分形概念

本章中已经对动力系统、湍流、降雨量过程和地震事件这些例子作了简短讨论, 将这些例子放到一个测度空间 $(\chi, \mathcal{B}(\chi), \mu)$ 中, 在例 1.2.1 的 Cantor 测度中, $\chi = [0,1]$, 测度 μ 的支撑是 Cantor 集, 这个 Cantor 集的分形维数为 $\log_3 2$(盒维数和 Hausdorff 维数均是). 在动力系统情形, $\chi \subset \mathbb{R}^d$ 是相空间, 测度 $\mu(A)$ 可以看成包含任何给定时刻的轨道的集合 $A \in \chi$ 的概率. 在关于湍流的讨论中, μ 是能量消耗的测度, 在自然中是随机的.

在每个例子中, 测度 μ 的支撑可能非常不规则, Lebesgue 测度可能为零, 测度本身也可以是多种多样的而且有很多不同阶的奇异性, 这样的测度没有密度, 如何来刻画这种测度呢? 这些测度中许多有类似重分形的性质, 本章中直观地介绍了重分形, 许多概念都是相当粗略地定义的, 在后面的章节中将给出更精确的定义.

考虑用宽为 δ_n 的 d-维盒子来覆盖 χ 的格点覆盖, 记 $B_{\delta_n}(x)$ 是包含 x 点的盒子, 当 $n \to \infty$ 时, $\delta_n \to 0$. 考虑另一个映射

$$U_n : \chi \to \mathbb{R},$$

其中当 $\mu[B_{\delta_n}(x)] > 0$ 时, $U_n(x) = -\log \mu[B_{\delta_n}(x)]$. 记 $Y_n(x)$ 是 $U_n(x)$ 一个重新刻画, 即

$$Y_n(x) = \frac{U_n(x)}{-\log \delta_n}.$$

变量 Y_n 描述了测度 μ 的局部或单点行为.

在大多数情形, Y_n 的极限概率分布是平凡的, 也即是说, 存在一个数 y_0, 使得当 $y \neq y_0$ 时,

$$\lim_{n\to\infty} \Pr\{Y_n = y\} = 0.$$

另外一种刻画 Y_n 的方式是描述在该点的概率趋于 0 的速率, 也即是像例 1.2.1 中 Y_n 有某一给定速率的盒子的个数, 这就是所谓的重分形谱 $\tilde{f}(y)$, 可以表示为

$$\tilde{f}(y) = \lim_{\varepsilon\to 0}\lim_{n\to\infty} \frac{\log\{\text{在第 } n \text{ 级具有正测度且满足 } |Y_n - y| < \varepsilon \text{的盒子数}\}}{-\log \delta_n}.$$

其中 $\tilde{f}(y)$ 与计盒维数类似, 但也并不相同, 这是因为 $n \to \infty$ 时被宽度减小的盒子覆盖的集合的特征也是在不断地变化着, 记 N_n 为第 n 阶有正测度的盒子数, 则

$$\tilde{f}(y) = \lim_{n\to\infty} \frac{\log N_n}{-\log \delta_n} + \lim_{\varepsilon\to 0}\lim_{n\to\infty} \frac{\log \Pr\{|Y_n - y| < \varepsilon\}}{-\log \delta_n}. \tag{1.8}$$

第一项是 μ 的支撑的计盒维数, 假设 $y \neq y_0$, 第二项是 Y_n 趋于 0 的概率函数的幂定律, 这一项是大偏差理论所关心的焦点, 也经常被看做是熵函数.

现在考虑 $U_n(x)$ 的一个重整累积生成函数, 记为 $\tilde{\theta}(q)$, 即

$$\tilde{\theta}(q) = \lim_{n\to\infty} \frac{\log E[\exp(-(q-1)U_n(X))]}{\log \delta_n} = \lim_{n\to\infty} \frac{\log E[\mu^{q-1}[B_{\delta_n}(X)]]}{\log \delta_n}.$$

这里, 对随机变量 X 的期望是关于给定的概率分布 μ 的. 在 §1.4 和 §1.5 的情形, μ 是个随机变量, 但函数形式在本质上是一样的, 需要注意的是 $\tilde{\theta}(q)$ 将有两种不同的情形出现. 首先, Rényi 维数是 $\tilde{D}_q = \tilde{\theta}(q)/(q-1)$, 在方程 (1.5) 和 (1.6) 中均出现过, 分别对应于简单尺度和更复杂尺度的情形. 在简单尺度情形, 所有的 Rényi 维数都是一样的, 但在复杂情形下它们是不同的. 回忆一下在例 1.2.1 中 Cantor 测度的支撑的维数是不随概率的分配的不同而变化的, 但是 Rényi 维数在 Cantor 集中是随着概率分配的不同而变化的. 第二个需要说明的情况是重整累积生成函数是一种全局平均, 且出现在大偏差理论中, 这些将在本书的第二部分中用到.

分别通过 $\tilde{\theta}(q)$ 和 $\tilde{D}_q$ 给出的全局平均或 Rényi 维数经常通过一个 Legendre 变换与重分形谱 $\tilde{f}(y)$ 联系起来, 我们感兴趣的是有这种联系的情形. 在这些情形, 将 μ 看做弱意义下的重分形测度. 形式的定义将在第 2 章给出, 重整累积生成函数和在大偏差理论中被称为熵函数 (也就是方程 (1.8) 的右边的最后一项) 之间也有同样的关系. 因此将使用这一理论来确定弱意义下测度 μ 成为重分形测度的必要条件.

对于一些测度, $\tilde{f}(y)$ 有一个更强的解释, 也就是, $\tilde{f}(y) = \dim_H \tilde{F}(y)$, 其中

$$\tilde{F}(y) = \{x : \lim_{n\to\infty} Y_n(x) = y\} = \left\{x : \lim_{n\to\infty} \frac{\log \mu[B_\delta(x)]}{\log \delta_n} = y\right\},$$

$\dim_H$ 是 Hausdorff 维数. 因此在这种情形, 在 §1.2 中默认的划分是 $\tilde{F}(y)$, 分形维数是 Hausdorff 维数. Mandelbrot (1989) 写到 (他用记号 $f(\alpha)$ 代替 $\tilde{f}(y)$): *重分形测度可以表示为连续的无穷个加数的并, 每个加数是一个无穷小的"一致分形测度", 它被 α 的单个值所刻画, 被分形维数为 $f(\alpha)$ 的分形集所支撑, 不同的 α 所对应的集合是错综复杂的.*

通过格点覆盖来描述测度 μ 不是唯一可行的方法, 也可以通过下面的量来描述局部行为

$$\frac{\log \mu[S_\delta(x)]}{\log \delta},$$

其中 $S_\delta(x)$ 是中心在 x 半径为 δ 的一个球. 也可以给出与全局平均 $\tilde{\theta}(q)$ 相类似的定义, 将这种框架称为点中心构造.

上面描述的是 μ 为概率测度的情形, 另一种可能是 μ 为通过某种过程构造的随机测度, 这种情形出现在湍流模型中, 在 §1.4 中进行了讨论.

上面已经简单地描述了刻画重分形行为的两种方法：一种是一个基于格点的构造, 另一种是基于点中心的构造, 这两种方法都与从局部行为到全局平均相关. 进一步, 分形维数有弱的或强的意义下的解释, 在文献中, 这些通常相应地被说成是粗糙的和精细的 (Falconer, 1997), 重分形的定义及构造方法将在第 2 章给出.

1.9 全书概述

本书分为三个部分.

第一部分 —— 引言和预备知识

本章已经描述了重分形测度 μ 的各种各样的特征, 测度可以像例 1.2.1 中对 Cantor 集那样以确定的方式构造, 也可以像 §1.4 和 §1.5 中是随机的. 它的支撑可能是个分形集, 其维数不依赖于测度的分配方式, 能够通过观察测度的局部或全局行为来描述其分布, 在某些情况下, Legendre 变换把它们联系起来. 同样的关系在大偏差理论中也是成立的 (也就是方程 1.8 中的 $\tilde{\theta}(q)$ 和熵的关系), Rényi 维数基于整体 (平均) 的行为, 在实际研究中对分形维数的估计一般是估计其 Rényi 维数, 这种估计可以用来估计重分形谱 $\tilde{f}(y)$, 至少从理论上来说是可以的.

在本章中, 各种各样的技巧都得到了运用, 但有时是相当粗糙的. 在第 2 章中, 弱和强意义下的重分形测度将通过格点和点中心构造的方式给出定义. 在多项测度情形, Cantor 测度是一种特殊情形, 重整累积生成函数 $\tilde{\theta}(q)$ 和重分形谱 $\tilde{f}(y)$ 之间的 Legendre 变换关系将通过 Lagrange 乘子法很容易地给出证明, 这个例子将在第 3 章中讨论.

第二部分 —— 大偏差下的重分形公式

在第二部分, 将分别在第 4 和第 5 章使用大偏差中的 Gärtner-Eills 定理作为基础来提供与格点和点中心构造平行的理论, 大偏差中的 Gärtner-Eills 定理及其他相关结果在附录 B 中作了介绍. 注意到, 在基于格点的构造中在 $\theta(q)$ 和 $f(y)$ 上使用 $\sim$ 符号, 而在点中心构造中上面没有 $\sim$ 符号. 我们想寻找测度的全局和局部行为能够通过 Legendre 变换联系起来的一般性条件, 也对重分形谱能够解释为 Hausdorff 维数的条件感兴趣.

给定基于格点和点中心的构造, 在什么条件下, 这两种构造的全局和局部行为是相容的? 这个问题将在第 5 章的最后讨论. 第 6 章是第二部分的结论, 这一章将总结其他构造方法, 有些方法不能很好地用大偏差的结构来描述.

第三部分 ——Rényi 维数的估计

第三部分的重点是 Rényi 维数的估计, 在前面的讨论中, 已经涉及基于格点和点中心的两种构造方法. 在估计过程中这两种方法是有区别的, 如果用基于格点的情形来估计 Rényi 维数, 可以相应地通过盒子的格点系统来覆盖、计数及取平均. 在点中心情形. 我们分析相互点的距离. 事实上, 在点中心情形, 在重整累积生成函数中出现的期望项 $E[\mu^{q-1}[S_\delta(X)]]$ 是当 $q=2,3,\cdots$ 时一个 q 阶相互点距离 (记为 Y) 的概率分布函数, 也即是 $\Pr\{Y\leqslant y\}=F_Y(y)E[\mu^{q-1}[S_y(X)]]$. q 阶的相互点距离 (也就是 Y) 将在第 2 章中定义. 因此 Rényi 维数本质上是概率函数 $F_Y(y)$ 的幂指数 (假设这样的幂指数存在), 也即是 $F_Y(y)\sim y^{\theta(q)}$. 在这点上, 问题看起来相对容易, 对不同的 q 值, 可以画出许多 q 阶点间距的一个样本, 从而估计 $\theta(q)$, 因此可以部分地重新构造重分形谱. 不幸的是, 这个问题并不是十分简单, 这是因为在估计 $\theta(q)$ 中偏差有各式各样的形式.

在第 7 章中, 将定义相关积分, 这个与第 2 章中的 Rényi 维数的定义有关, 在对关联指数的估计中, 相关积分的内在特征可以导致偏差, 这些将在第 7 章中讨论. 当测度支撑在某些自相似集上时, 这些相当容易. 这导致函数 $F_Y(y)$ 有一个振动行为, 这种行为在对数尺度上是周期的, 这就意味着 $F_Y(y)$ 在平均意义下仅有幂定律行为.

我们已经广泛地运用一个修改的 Hill 估计函数来估计 Rényi 维数, 这些将和一些其他的估计方法一起在第 8 章讨论. 第 9 章描述偏差的各种外部资源, 也就是说, 不是内在于关联积分本身, 而是可能归因于抽样策略和其他缺失处理的数据, 这些误差跟样本分析中的非样本误差类似. 这里三个主要问题是噪声、数据中的误差、数据的舍入和边界效应, 边界效应通常发生在基于相互距离的估计方法中, 数据的舍入和噪声有模糊精细尺度信息的效果. 由于维数是一个当相互距离非常小时的极限概念, 故偏差的这些形式也是非常严格的.

在第 10 章, Rényi 维数将通过使用从各种统计的和数学模型仿真的数据来估计, 在其中的一些模型中, 维数可以被解析地计算. 在其他模型中, 很多研究者给出了估计, 关于精确值是多少也有一些共识, 甚至在这些分析中, 偏差的内在的和外在的形式都很明显, 偏差的这些形式也需要从估计函数自身来理清, 通过至少部分性质被了解的模型来帮助理解一般的估计问题.

第 11 章为第三部分的结论, 本章中使用发生在新西兰和日本的地震震源位置估计了 Rényi 维数, 这些数据是有趣的, 不仅因为从地震预测的观点来说如此, 还因为它包含了本书第三部分所讨论的偏差的很多形式.

第 2 章　重分形公式

2.1 引　　言

本章给出 Rényi 维数的定义以及分别给出格和点中心的重分形谱的定义. 给出重分形公式的定义, 并且复习相关文献的已知结果.

设 $\mathcal{X}$ 为 $\mathbb{R}^d$ 的一个 Borel 子集, $\mathcal{B}(\mathcal{X})$ 是 $\mathcal{X}$ 中所有 Borel 集组成的类. 考虑概率空间 $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu)$, 其中无原子的概率测度 μ 可支撑在 $\mathcal{X}$ 的维数低于 d 的子空间上. 例如, 在 Lorenz 吸引子的情形下 (见 §1.3.3), $d = 3$, 然而出现了吸引子集中在两个圆盘上的情形 (见图 1.5), 实证研究表明吸引子的维数严格大于 2. 测度 $\mu(A)$ 给出了在任意时间集合 A 中出现轨道的概率.

注意到, 如果需要描绘基于时间变化观察到的轨道路径的空间特征, 测度 μ 的不变性及遍历性的假设是必需的. 不变性是指测度 μ 在方程 (1.1) 的变换下保持不变, 即 $\mu = \mu T_\xi^{-1}$, 遍历的意思是对 μ-几乎所有的 x, 时间平均 (即反复迭代 T_ξ^{-1} 后的平均) 等于空间平均. 关于这些思想的进一步详细讨论可在 Walters (1982) 及 Billingsley (1965) 的文章中找到. 考虑如下两种重分形结构.

1) $\mathcal{X}$ 被一系列 d 维盒子的格子覆盖, 其中盒子的宽度 δ_n 随着 $n \to \infty$ 递减. 称这种情况为格情形.

2) 对所有 x, $\mu(S_\delta(x)) > 0$, 其中 $S_\delta(x)$ 是半径为 δ 中心在 x 的闭球. 称此为点中心情形.

与格情形有关及与点中心情形易混淆的函数会受波形符的影响.

我们希望在上述两种情况下描述整体和局部行为, 整体行为的概念基于匈牙利数学家 Alfred Rényi 关于信息论的工作. 这个工作也是 §2.3 和 §2.4 中导出的 Rényi 维数的基础. 关于它的简要复习在 §2.2 中, 局部行为用重分形谱描述, 这个定义见 §2.5.

重分形公式涉及到用 Legendre 变换公式表达的整体 (Rényi 维数) 和局部行为 (重分形谱) 之间的关系. 称这个为弱情形. 强情形即重分形谱被解释为一族 Hausdorff 维数. 强和弱两种情形下的重分形公式的定义见 §2.5.

我们的定义不是描述重分形公式的唯一方式. 存在多种定义, 每一种都有其内在差异和弱点. 在 §2.6 和 §2.7 分别回顾这些问题和另一些涉及格和点中心结构的结果.

注意到, 在 §1.4 中讨论湍流时提到的随机测度与本章讨论的概率测度是不同的. 在本章及第 4、5 章中, 总假设是概率测度. 关于随机测度的初次讨论见 §1.4, 在第 6 章会再次提到.

2.2 广义 Rényi 维数的发展历史

Rényi 维数源于信息论. 这个理论的出现涉及信息的传输, 特别是信息的二进制表示的长度. 设集合 E 有 n 个元素, 如果 $n=2^N, N\in\mathbb{Z}^+$, 任意一个元素可用一个 N 位二进制数标记. 因此 Hartly 将 $\log_2 n$ 定义为描述 E 必需的信息.

设 $E=\bigcup\limits_{k=1}^{b}E_k$, 其中 $E_1,E_2,\cdots,E_b$ 为互不相交的有限集. 做如下实验: 按概率 p_k 独立地随机地分配 n 个元素到 b 个子集 $E_k(k=1,\cdots,b)$, 关于概率分布 $P=(p_1,\cdots,p_b)$ 的这个试验产生的信息量为

$$H_1(P)=-\sum_{k=1}^{b}p_k\log_2 p_k,$$

这就是著名的 Shannon(香农) 公式.

Rényi (1965) 证明 Shannon 公式能按如下方式推导. 假设有一个令人感兴趣的特殊元素, 并且不知道该元是这 n 个元素中的哪一个. 完成上述系列试验, 即按概率 p_k 将 n 个元素独立随机分配到 b 个子集 $E_k(k=1,\cdots,b)$. 每次试验后, 我们仅知道每个元素被分配到哪个子集及哪个子集包含未知的令人感兴趣的元素. 第一次试验产生分划 Δ_1, 第二次实验产生分划 Δ_2, 以此类推. 记 $\Delta^{(m)}$ 为前 m 个独立的试验产生的 m 个分划的叉积. 每一个元素可认为在 $\Delta^{(m)}$ 中有一个长为 m 的路径. 当未知元的路径相对其他所有元是唯一的时, 未知元被唯一确定. 设 P_{nm}^* 为这个未知元素在 m 次试验后能被唯一确定的概率, 即它的路径是唯一的. 定义

$$e_1(n,\varepsilon)=\min\{m:P_{nm}^*\geqslant 1-\varepsilon\},$$

其中 $0<\varepsilon<1$, 则 Rényi 证明:

$$\lim_{n\to\infty}\frac{\log_2 n}{e_1(n,\varepsilon)}=H_1(P).$$

一个启发式的解释如下. 描述 E 所需要的信息总量为 $\log_2 n$. 每次实验产生 $H_1(P)$ 信息, 当 n 足够大时, 需要的实验次数近似为 $\dfrac{\log_2 n}{H_1(P)}$.

一个勿需证明的方法也可以导出 $H_1(P)$ 的公式, 并且是唯一的具有所需要性质的函数. 这个性质即可加性. 考虑两个不同的概率分布及相应的两种分划:

$$P=(p_1,\cdots,p_b)\text{ 及 }Q=(q_1,\cdots,q_a).$$

设 $P*Q$ 是项为 $p_j q_i (i=1,\cdots,a; j=1,\cdots,b)$ 的概率分布, 则 H_1 满足可加性, 即

$$H_1(P*Q) = H_1(P) + H_1(Q). \tag{2.1}$$

由此, Rényi(1965) 把有关信息的概念拓展到了高阶. 设 $P_{nm}^{(q)}$ 表示至多 q 个元素在 $\Delta^{(m)}$ 中路径相同的概率, 令

$$e_q(n,\varepsilon) = \min\{m : P_{nm}^{(q)} \geqslant 1-\varepsilon\}, \quad q=2,3,\cdots,$$

则 Rényi 证明:

$$\lim_{n\to\infty} \frac{\log_2 n}{e_q(n,\varepsilon)} = \left(1-\frac{1}{q}\right) H_q(P), \quad q=2,3,\cdots,$$

其中

$$H_q(P) = \frac{-1}{q-1} \log_2 \sum_{k=1}^{b} p_k^q.$$

进一步, $\lim\limits_{q\to 1} H_q(P) = H_1(P)$, 并且 $H_q(P)$ 满足可加性 (2.1). 如同 $H_1(P)$ 一样, $H_q(P)$ 的函数表达式可由实际的角度或公理化方法导出 (Rényi, 1965, 1970). 对 $q<0$ 的情形的进一步讨论见 Rényi (1970,581 页).

1959 年 Rényi 导出维数的思想如下. 考虑一个以概率 p_k 取可数个值 x_k 的随机变量 X, 概率 $p_k = \Pr\{X = x_k\}$. 可定义包含 X 的值的信息为

$$H_1(P) = -\sum_{k=1}^{\infty} p_k \log_2 p_k, \quad q=1,$$

$$H_q(P) = \frac{-1}{q-1} \log_2 \sum_{k=1}^{\infty} p_k^q, \quad q \neq 1.$$

若 X 的分布是连续的且 X 的值是确定的, 则信息的总和是无穷的.

1959 年 Rényi 对离散随机变量 $X_n = [nx]/n$, 其中 $[x]$ 表示 x 的整数部分, 描述了 X 以及当 $n\to\infty$ 时的行为. 令 $P_n = (p_1, p_2, \cdots)$ 为 X_n 的概率分布, 定义分布的维数为

$$d_q = \lim_{n\to\infty} \frac{H_q(P_n)}{\log_2 n}.$$

这个量告诉我们 X 的信息趋于无穷的速度. 对该维数的进一步讨论见 Rényi (1970, 588 页).

2.2.1 例子

设 X 是一个随机变量, 它从如例 1.2.1 定义的 Cantor 测度所给的分布中抽样. 考虑像例 1.3.1 同样的情形, $\omega_1, \omega_2, \cdots$ 是单边无穷序列, ω_i 取 0 和 2 的概率分别为 p_0 和 p_2, 考虑随机变量 X, 记三进制表示为 $X = (0.\omega_1\omega_2\omega_3\cdots)_3$. 令

$$\begin{aligned} X_1 &= (0.\omega_1)_3, \\ X_2 &= (0.\omega_1\omega_2)_3, \\ &\vdots \\ X_n &= (0.\omega_1\omega_2\cdots\omega_n)_3 = \frac{[3^n X]}{3^n}. \end{aligned}$$

在每一步, X 的三进表示会增加一位数字. 随着 n 的增加, 从 Cantor 测度 μ 中积累的信息以怎样的速度增加呢?

设 p_n 为离散随机变量 X_n 的概率分布, 取 2^n 种可能的值 $x_k, k = 1, \cdots, 2^n$. 在 $q \neq 1$ 的情形,

$$\begin{aligned} H_q(P_n) &= \frac{-1}{q-1} \log_2 \sum_{k=1}^{2^n} (\Pr\{X_n = x_k\})^q \\ &= \frac{-1}{q-1} \log_2 (p_0^q + p_2^q)^n \\ &= \frac{-n}{q-1} \log_2 (p_0^q + p_2^q). \end{aligned}$$

因此, 对 $q \neq 1$ 的 Rényi 维数是

$$d_q = \lim_{n\to\infty} \frac{H_q(P_n)}{\log_2(3^n)} = \frac{-1}{q-1} \log_3 (p_0^q + p_2^q).$$

类似地, 当 $q = 1$ 时,

$$d_1 = \lim_{n\to\infty} \frac{H_1(P_n)}{\log_2(3^n)} = -(p_0 \log_3 p_0 + p_2 \log_3 p_2).$$

在例 1.2.1 中, 用维数描述了 Cantor 测度的支撑的大小, 这个刻画了测度被分配到其支撑的方式. 通过 Rényi 维数, 能刻画测度被分配的方式的特征. 在下面几节中, 定义更一般情形的 Rényi 维数.

2.3 广义 Rényi 格维数

对 μ 的支撑, 考虑它的一个宽为 δ_n 的 d-维盒子格覆盖, 盒子左闭右开, 在每个盒右端实现连接. k 阶盒用 $B_{\delta_n}(k)$ 表示, 其中 $k \in K_n$, $K_n = \{k : \mu(B_{\delta_n}(k)) > 0\}$. 我们估计连续的格覆盖, 考虑当 $n \to \infty$ 时的 $\delta_n \to 0$ 的序列 $\{\delta_n\}$.

根据 Rényi 的有关信息论的文章, 可考虑如下情况. 设 $\mathcal{F}_n$ 表示由盒子 $B_{\delta_n}(k)$ 生成的 σ-域, 其中 $k \in K_n$. 我们的兴趣是概率空间 $(\mathcal{X}, \mathcal{F}_n, P_n)$, 其中 $P_n(A) = \mu(A), A \in \mathcal{F}_n$, 对所有可能的求和与对所有格盒子求和一致. 因此, 对 $q \neq 1$,

$$H_q(P_n) = \frac{-1}{q-1}\log_2 \sum_{k \in k_n} \mu^q[B_{\delta_n}(k)].$$

则维数用盒子的宽度的标度定义.

2.3.1 定义

设 $K_n = \{k : \mu(B_{\delta_n}(k)) > 0\}$, 若下列极限存在, 定义 $\widetilde{\theta}(q)$ 为

$$\widetilde{\theta}(q) = \lim_{n \to \infty} \frac{\log \sum\limits_{k \in K_n} \mu^q[B_{\delta_n}(k)]}{\log \delta_n}, \quad -\infty < q < \infty. \tag{2.2}$$

注意, 这里允许 $-\infty$ 作为其极限值.

用 $k(x)$ 表示含 x 的盒子的标志, 则 $B_{\delta_n}(k(x))$ 是含 x 的盒子. 为方便用 $B_{\delta_n}(k)$ 表示 k-级盒子, 用 $B_{\delta_n}(x)$ 表示含 x 的盒子. 可以看到方程 (2.2) 的和正好为 $E[\mu^{q-1}[B_{\delta_n}(X)]]$, 其中期望是关于概率测度 μ 的.

2.3.2 定义

若 $q=1$ 时下列极限存在, 且 $q \neq 1$ 时 $\widetilde{\theta}(q)$ 存在, 则定义广义 Rényi 格维数 $\widetilde{D}_q$ 为

$$\widetilde{D}_q = \begin{cases} \lim\limits_{n \to \infty} \dfrac{\sum\limits_{k \in K_n} \mu[B_{\delta_n}(k)] \log \mu[B_{\delta_n}(k)]}{\log \delta_n}, & q = 1, \\ \dfrac{\widetilde{\theta}(q)}{q-1}, & q \neq 1. \end{cases} \tag{2.3}$$

注意到 $\widetilde{\theta}(1) = 0$. 进一步, 若 μ 的支撑的盒维数存在 (定义 A.3.1), 则它为 $\widetilde{D}_0$, 并且

$$\widetilde{D}_0 = \lim_{n \to \infty} \frac{\log \# K_n}{-\log \delta_n},$$

其中 $\# K_n$ 表示 K_n 的个数, 即使 μ 为正的盒子的个数.

至此, 将 $\widetilde{D}_q$ 作为一个整体均值或信息论中的信息测度进行了解释. 当然, 还有第三种解释, 在第 4 章中被解释为一个按比例累积生成的函数.

2.3.3 定理 (Beck, 1990)

下面的性质对任何概率测度成立:

1) $\widetilde{D}_r \leqslant \widetilde{D}_q$, 对 $r > q$ 及 $q, r \in \mathbb{R}$.

2) $\frac{r}{r-1}\frac{q-1}{q}\widetilde{D}_q \leqslant \widetilde{D}_r \leqslant \widetilde{D}_q$, 对 $r>q>1$ 或 $0>r>q$.

Beck (1990) 称一个测度 μ 具有最低均匀度, 若

$$\frac{q-1}{q}\widetilde{D_q} = \frac{r-1}{r}\widetilde{D}_r = 常数,$$

称测度 μ 具有最高均匀度, 若 $\widetilde{D}_q = \widetilde{D}_r = $ 常数.

2.3.4 例子

考虑例 1.2.1 和例 2.2.1 中的 Cantor 测度. 当测度不是均匀分配到 Cantor 集上时, 例如在例 1.2.1 中 $p_0 = 1-p_2 = \frac{1}{3}$, 可得到 $\widetilde{\theta}(q) = -\log_3(p_0^q + p_2^q)$ 及 $\widetilde{D}_q = \frac{\widetilde{\theta}(q)}{q-1}$.

当测度均匀分配到 Cantor 集上时, 即 $p_0 = 1-p_2 = \frac{1}{2}$, 则 $\widetilde{\theta}(q) = (q-1)\log_3 2$ 及 $\widetilde{D}_q = \log_3 2$ 对所有 q 成立. 这就是定理 2.3.3 中提到的最高均匀度, $\widetilde{D_q}$ 是 Cantor 集的维数.

2.4 广义 Rényi 点中心维数

我们研究的第二种重分形公式是基于中心在测度 μ 的支撑上的球覆盖. 考虑这种结构的部分原因是为了采用更有效的算法来估计分形维数. 这里简要介绍这种算法的轮廓, 更详细的讨论见本书第三部分.

如果要估计观察过程的盒维数, 可用宽为 δ_n 的盒子形成的格序列覆盖, 然后数覆盖它的盒子数 N_n. 对递减的序列 $\{\delta_n\}$ 重复上述做法. 绘制 $\frac{\log N_n}{\log \delta_n}$ 的图. 当 δ_n 充分小时, 图将是一条直线, 其斜率是盒维数的一个估计. 当然, 计算过程会花大量工夫, 特别是 δ_n 非常小的时候, 因为大部分球在研究过程中都未观察到. 注意到, 通过观察的数据来估计 μ 的支撑的维数是非常困难的, 因为支撑的部分可能在研究过程中很少被观察到.

Grassberger 和 Procaccia (1983a,b,c) 提出一个可供选择的方法. 给一列随机地址 $X_1, X_2, \cdots$, 估计内点距离的概率分布, 然后估计这个概率分布的幂指数 (分形维数). 即取许多独立的样本点对, 估计 δ 的函数 $\Pr\{\|X_1 - X_2\| \leqslant \delta\}$, 在盒子计数情形绘制图 $\frac{\log \Pr\{\|X_1 - X_2\| \leqslant \delta\}}{\log \delta}$ 并估计直线的斜率. 这就是后面提到的相关维数. 注意到

$$\Pr\{\|X_1 - X_2\| \leqslant \delta\} = \int \mu[S_\delta(x)]\mu(dx) = E[\mu[S_\delta(x)]], \tag{2.4}$$

其中 $S_\delta(x)$ 是中心为 x 半径为 δ 的闭球, 期望是关于概率测度 μ 的. 如同格情况一样, 估计一阶矩的幂指数, 然而并非所有的球都像格情况下的盒子一样具有相同的作用, 这是根据过程经过空间部分的概率给出的粗略估计.

一般地, 考虑 $(q-1)$ 阶矩

$$E[\mu^{q-1}[S_\delta(x)]] = \int \mu^{q-1}[S_\delta(x)]\mu(dx).$$

对给定值 q, Rényi 指出点中心维数描述了当 $\delta \to 0$ 时 $\{E[\mu^{q-1}[S_\delta(x)]]\}^{q-1}$ 的幂行为. Cutler 在 1991 年通过 L^q 范数性质刻画了这些维数.

2.4.1 定义

设 $-\infty < q < \infty$, 且 $\mathcal{X}_\delta = \{x \in \mathcal{X} : \mu[S_\delta(x)] > 0\}$, 如果下列极限存在, 点中心相关指数定义为

$$\theta(q) = \lim_{\delta \to 0} \frac{\log \int_{\mathcal{X}_\delta} \mu^{q-1}[S_\delta(x)]\mu(dx)}{\log \delta}, \tag{2.5}$$

注意到 $\theta(q) = -\infty$ 是允许的. 当 (2.5) 中的极限不存在时, 考虑上下极限. 注意到, 若 $q < 0, \theta(q) = -\infty$, 则当 $\delta \to 0$ 时 $E[\mu^q[S_\delta(x)]] \to \infty$ 且它趋于 ∞ 的速度比指数率更快.

2.4.2 定义

广义 Rényi 点中心维数记为 D_q(对 $q=1$ 下列极限存在时, 对 $q \neq 1$, $\theta(q)$ 存在时), 其定义如下

$$D_q = \begin{cases} \dfrac{\theta(q)}{q-1}, & q \neq 1, \\ \lim\limits_{\delta \to 0} \dfrac{\int_{\mathcal{X}_\delta} \log \mu[S_\delta(x)]\mu(dx)}{\log \delta}, & q = 1. \end{cases} \tag{2.6}$$

相对计算 $\widetilde{D}_0$, Grassberger 和 Procaccia (1983a,b,c) 计算 D_2 采用了更高效的方法, 他们介绍了在观察过程中发现的与众不同的特征. μ 的支撑的盒维数反映了几何维数或支撑的大小, 而 Rényi 维数 D_q(及 $\widetilde{D}_q, q \neq 0$) 描述了测度的非均匀性.

2.4.3 注记

如果 μ 为定义在其支撑上, 具有相对均匀测度的有界的 Radon-Nikodym 导数 (即 $\mu(E) = \int_E g(x)dx, |g(x)| \leqslant K$), 则对于所有 $q > 0$, D_q 为常数. 然而, 若 μ 有奇点, Rényi 维数将不一致, 并刻画了奇点的特征. 在这种情况下, $D_q(q \neq 1)$ 为 q 的增函数.

这个注与格情形下的定理 2.3.3 类似.

2.4.4 例子

设 μ 是 [0,1] 上的均匀测度. 则对于 $0<\delta<\dfrac{1}{2}$,

$$\mu[S_\delta(x)]=\begin{cases}\delta+x, & x<\delta,\\ 2\delta, & \delta\leqslant x\leqslant 1-\delta,\\ \delta+1-x, & x>1-\delta.\end{cases}$$

因此, 对于 $q\neq 0,\ 0<\delta<\dfrac{1}{2}$,

$$\begin{aligned}&E[\mu^{q-1}[S_\delta(x)]]\\ =&\int_0^1\mu^{q-1}[S_\delta(x)]\mu(dx)\\ =&\int_0^\delta(\delta+x)^{q-1}dx+\int_\delta^{1-\delta}(2\delta)^{q-1}dx+\int_{1-\delta}^1(\delta+1-x)^{q-1}dx\\ =&(2\delta)^{q-1}\left[1+\frac{2\delta}{q}(2-q-2^{1-q})\right].\end{aligned}$$

注意到, 若 $q<0, \lim_{\delta\to 0}E[\mu^{q-1}[S_\delta(x)]]=\infty$. $\theta(q)$ 描述了当 $\delta\to 0$ 时 $E[\mu^q[S_\delta(x)]]$ 的指数行为. 可以看到 $\theta(q)=\lim_{\delta\to 0}\theta_\delta(q)=q-1$, 所以 $D_q=1$ 对所有 $q\neq 0$ 成立, 与注记 2.4.3 一致.

由等式 (2.4) 确定的与内点距离相关的概率分布的概念可以拓展到 q 个独立的点. 这种关系是本书第三部分中用来估计 D_q 的基础.

2.4.5 q 阶点间距的分布定理

设 $X_1, X_2, \cdots, X_q$ 为关于概率分布 μ 独立的随机变量, 定义 Y 为

$$Y=\max\{\|X_1-X_q\|,\|X_2-X_q\|,\cdots,\|X_{q-1}-X_q\|\}.$$

则对于 $q=2,3,4,\cdots$,

$$\int\mu[S_\delta(x)]^{q-1}\mu(dx)=\Pr(Y\leqslant\delta).$$

证明 设 $1(A)$ 为 A 的特征函数, 则

$$\mu[S_\delta(x)]=\int 1(\|x_1-x\|\leqslant\delta)\mu(dx_1)=\Pr(\|X_1-x\|\leqslant\delta).$$

从而

$$\mu^{q-1}[S_\delta(x)]=\int\cdots\int \mathbf{1}(\|x_1-x\|\leqslant\delta)\cdots\int \mathbf{1}(\|x_{q-1}-x\|\leqslant\delta)\mu d(x_1)\cdots\mu d(x_{q-1})$$
$$=\int\cdots\int \mathbf{1}(\max\{\|x_1-x\|,\cdots,\|x_{q-1}-x\|\}\leqslant\delta)\mu(dx_1)\cdots\mu(dx_{q-1}),$$

故

$$\int \mu^{q-1}[S_\delta(x)]\mu(dx)$$
$$=\iint\cdots\int \mathbf{1}(\max\{\|x_1-x\|,\cdots,\|x_{q-1}-x\|\}\leqslant\delta)\mu(dx_1)\cdots\mu(dx_{q-1})\mu d(x_q)$$
$$=\Pr(Y\leqslant\delta).$$

□

2.5 重分形谱和重分形公式

这里介绍测度 μ 的局部行为, 即在单独的格子或球上的行为.

2.5.1 定义

设

$$K_n(y,\varepsilon)=\left\{k: y-\varepsilon<\frac{\log\mu[B_{\delta_n}(k)]}{\log\delta_n}\leqslant y+\varepsilon\right\}.$$

格情形的重分形谱, 记为 $\widetilde{f}(y)$, 对 $y>0$ 定义

$$\widetilde{f}(y)=\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{\log\#K_n(y,\varepsilon)}{-\log\delta_n}. \tag{2.7}$$

当 $\#K_n(y,\varepsilon)=0$ 时, 允许 $\widetilde{f}(y)=-\infty$.

函数 $\widetilde{f}(y)$ 可以有不同的表示. 令 $B_{\delta_n}(k(x))$ 表示包含 x 的盒子, 即 x 包含在第 k 个盒子中. 为避免记号过于繁杂, 如果在上下文中不会和 $B_{\delta_n}(k)$(这里 $k\in K_n$) 相混淆, 就直接用 $B_{\delta_n}(x)$ 表示 $B_{\delta_n}(k(x))$, 若 $k\in k_n$. 注意到对所有 $x'\in B_{\delta_n}(x)$, $\mu[B_{\delta_n}(x')]$ 为常数, 因此

$$\#K_n(y,\varepsilon)=\int_{\widetilde{F_n}(y,\varepsilon)}\frac{\mu(dx)}{\mu(B_{\delta_n}(x))},$$

其中

$$\widetilde{F_n}(y,\varepsilon)=\{x\in\mathcal{X}: y-\varepsilon<\frac{\log\mu[B_{\delta_n}(x)]}{\log\delta_n}\leqslant y+\varepsilon\}.$$

所以, $\widetilde{f}(y)$ 也可表示为

$$\widetilde{f}(y)=\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{-\log\delta_n}\log\int_{\widetilde{F_n}(y,\varepsilon)}\frac{\mu(dx)}{\mu(B_{\delta_n}(x))}. \tag{2.8}$$

等式 (2.8) 给出了点中心情形下重分形谱的类似式 (2.7) 的公式.

2.5.2　定义

令

$$F_\delta(y,\varepsilon)=\left\{x\in\mathcal{X}:y-\varepsilon<\frac{\log\mu[S_\delta(x)]}{\log\delta}\leqslant y+\varepsilon\right\}$$

为点中心情形的重分形谱. 记为 $f(y)$, 对 $y>0$ 定义

$$f(y)=\lim_{\varepsilon\to0}\lim_{\delta\to0}\frac{1}{-\log\delta}\log\int_{F_\delta(y,\varepsilon)}\frac{\mu(dx)}{\mu(S_\delta(x))}. \tag{2.9}$$

为了方便, 规定 $\log 0=-\infty$, 从而当上式中积分为零时, 允许 $f(y)=-\infty$.

2.5.3　注记

注意到, 若对所有 $x\in F_\delta(y,\varepsilon),\mu[S_\delta(x)]\approx$ 常数, 等式 (2.8) 和 (2.9) 相似. 则

$$\int_{F_\delta(y,\varepsilon)}\frac{\mu(dx)}{\mu(S_\delta(x))}\approx\frac{\mu(F_\delta(x))}{\text{const}}.$$

它大体上等于覆盖 $F_\delta(y,\varepsilon)$ 所需的球的个数.

2.5.4　定义

1) 称测度 μ 为在弱条件下满足格形式的重分形测度, 如果 $\widetilde{\theta}(q)$ 和 $\widetilde{f}(y)$ 分别由等式 (2.2) 和 (2.7) 定义, 定义存在且满足 Legendre 变换对

$$\widetilde{\theta}(q)=\inf_y\{qy-\widetilde{f}(y)\},$$
$$\widetilde{f}(y)=\inf_q\{qy-\widetilde{\theta}(q)\}.$$

2) 称测度 μ 为在强条件下满足格形式的重分形测度, 如果它满足弱条件, 且对 $\widetilde{f}(y)>0$,

$$\dim_H\widetilde{F}(y)=\widetilde{f}(y).$$

其中, $\widetilde{F}(y)=\{x\in\mathcal{X}:\lim\limits_{n\to\infty}Y_n(x)=y\}$, $\dim_H$ 是 Hausdorff 维数, 且

$$Y_n(x)=\frac{\log\mu[B_{\delta_n}(x)]}{\log\delta_n}. \tag{2.10}$$

由于 $\widetilde{f}(y)$ 的定义涉及某些盒覆盖, 似乎认为 $\dim_B\widetilde{F}(y)=\widetilde{f}(y)$. 其实并不如此, 公式 (2.7) 的极限不仅决定了该集合的性质, 还决定了格覆盖的大小. 从 $\widetilde{F}(y)$ 的定义直接得到确定 $\dim_B\widetilde{F}(y)$ 的覆盖盒子的大小和数量. 另外, 通常集合 $\widetilde{F}(y)$ 是稠密的, 因此 $\dim_B\widetilde{F}(y)=\dim_B\operatorname{supp}(\mu)$.

注意到, 有时 $\widetilde{f}(y)$ 被称作粗线条的重分形谱, $\dim_H\widetilde{F}(y)$ 或 $\dim_P\widetilde{F}(y)$ 被称作精细的重分形谱, $\dim_H$ 和 $\dim_P$ 分别为 Hausdorff 维数 (定义 A.2.3) 和 Packing 维数 (定义 A.4.2).

2.5.5 定义

1) 称测度 μ 为在弱条件下满足点中心形式的重分形测度, 如果 $\theta(q)$ 和 $f(y)$ 分别由等式 (2.5) 和 (2.9) 定义, 定义有意义且满足 Legendre 变换对

$$\theta(q) = \inf_y \{qy - f(y)\},$$
$$f(y) = \inf_q \{qy - \theta(q)\}.$$

2) 称测度 μ 为在强条件下满足点中心形式的重分形测度, 如果它满足弱条件, 且对 $f(y) > 0$,

$$\dim_H F(y) = f(y).$$

其中, $F(y) = \{x \in \mathcal{X} : \lim\limits_{\delta\to 0} Y_\delta(x) = y\}$, $\dim_H$ 是 Hausdorff 维数, 且

$$Y_\delta(x) = \frac{\log \mu[S_\delta(x)]}{\log \delta}. \tag{2.11}$$

注意集合 $\widetilde{F_n}(y,\varepsilon)$ 和 $\widetilde{F}(y)$, 以及 $F_\delta(y,\varepsilon)$ 和 $F(y)$ 的区别, $\widetilde{F_n}(y,\varepsilon)$ 和 $F_\delta(y,\varepsilon)$ 反映了预极限下的局部行为, $\widetilde{F}(y)$ 和 $F(y)$ 反映了极限的局部行为.

Mandelbrot 和 Riedi 在近期的一些文章中, 介绍了测度 μ 的逆, 记为 μ^*. 并用 "逆方程" $f^*(y) = yf(1/y)$ 代替重分形谱. 更详细的讨论见 1997, 1998 年 Riedi 和 Mandelbrot 以及 1997 年 Mandelbrot 和 Riedi (1997) 的文章.

2.6 格点情形的基本结论的复习

在格点情形下 $\widetilde{f}(y)$ 和 $\widetilde{\theta}(q)$ 的定义中, 极限由预序列 $\{\delta_n\} \to 0, n \to \infty$ 决定. 但是, 对一般序列 $(\delta \to \infty)$ 极限可能不存在. 另一种情形像等式 (2.2) 那样 limsup 和 liminf 不相等. 在这一节, 将复习这两种情形下的结论.

2.6.1 替代的关系 (Falconer, 1990, §17.1)

Falconer 用不同方式定义 $\widetilde{\theta}(q)$ 为

$$\widetilde{\theta}(q) = \lim_{\delta\to 0} \frac{\log \sum\limits_{k\in K_n} \mu^q[B_\delta(k)]}{\log \delta}, \tag{2.12}$$

即假定这个极限对所有序列存在. 注意到, 关于 $\widetilde{\theta}(q)$ 的存在性及某些性质将在第 4 章中研究, 由此 $\widetilde{f}(y)$ 的行为可被推断. Falconer (1990) 从另一方面提出问题, 对 $0 \leqslant y < \infty$ 假设等式 (2.7) 中的 $\widetilde{f}(y)$ 在 $\delta \to 0$ 的情况下存在 (对所有序列), 于是 Falconer 可证明等式 (2.12) 中的极限 $\widetilde{\theta}(q)$ 存在, 且 $\widetilde{\theta}(q) = \inf\limits_y \{qy - \widetilde{f}(y)\}$.

某些情况下这个极限依赖于序列 δ_n, 如 Riedi 在 1995 年给出的下列例子.

2.6.2　例子 (Riedi, 1995, §4.1)

设 μ 为例 2.3.4 中的 Cantor 测度. 则对于所有 $q \in \mathbb{R}$, 当 $\delta_n = 3^{-n}$ 时 $\widetilde{\theta}(q) = -\log_3(p_0^q + p_2^q)$, 在这个条件下, Riedi 在 1995 年证明当 $q < 0$ 时,

$$\liminf_{\delta \to 0} \frac{\log \sum_k' \mu^q[B_\delta(k)]}{\log \delta} = -\infty, \tag{2.13}$$

求和取遍所有具有正测度的盒子.

过程如下: 对于所有 $n \in \mathbb{N}$, 选择 $h_n \in \mathbb{N}$, 使得 $p_2^{h_n} \leqslant \left(\frac{1}{2}3^{-n}\right)^n$, 并且 $h_n \geqslant n+1$. 令 $\delta_n = 3^{-n}(1-3^{h_n})$, 则形成一个在 [0,1] 上的 δ_n-网, 最后一个盒子记为 $B_{\delta_n}(*)$, 它与 [0,1] 的交很小, 即

$$B_{\delta_n}(*) \cap [0,1] = [1-3^{-h_n}, 1],$$

它是最后一个宽为 3^{-h_n} 的三进区间. 因为 $1-3^{-h_n} > \frac{1}{2}$, 所以

$$\mu(B_{\delta_n}(*)) = p_2^{h_n} \leqslant \left(\frac{3^{-n}}{2}\right)^n < \delta_n^n.$$

从而当 $q < 0$ 时, $\sum_k' \mu^q[B_{\delta_n}(k)] > \delta_n^{nq}$. 故 $\frac{\log \sum_k' \mu^q[B_{\delta_n}(k)]}{\log \delta_n} < nq$, 等式 (2.12) 成立. □

2.6.3　重叠的盒子 (Riedi, 1995)

Riedi 在 $\delta \to 0$ 的情形对所有序列定义一个重叠盒子系统, 这样就避免了例 2.6.2 中的困难. 记 δ-网中的第 k 个盒子为 $B_\delta(k)$, 设 $\delta \to 0$ 是连续的, 而不是像定义 2.3.1 那样离散的. 特别地,

$$B_\delta(k) = [k_1\delta, (k_1+1)\delta) \times \cdots \times [k_d\delta, (k_d+1)\delta),$$

其中 $(k_1, \cdots, k_d)$ 是第 k 个球的网格坐标. 令 $K_\delta = \{k : \mu[B_\delta(k)] > 0\}$ Riedi 在 1995 年定义了一列有同样的盒子数量 $\#K_\delta$, 但盒子宽扩大 3 倍的序列, 因而产生了重叠. 记第 k 个盒子为 $B_\delta(k)$, 其中

$$\overline{B}_\delta(k) = [(k_1-1)\delta, (k_1+2)\delta) \times \cdots \times [(k_d-1)\delta, (k_d+2)\delta),$$

$k \in K_\delta$, 即 k 的指标与 $B_\delta(k)$ 中的格子指标相同.

他定义 $\widetilde{\theta}$-函数如下

$$\widetilde{\theta}_R(q) = \liminf_{\delta \to 0} \frac{1}{\log \delta} \log \sum_{k \in K_\delta} \mu^q[\overline{B}_\delta(k)],$$

加下标 R 区分 $\theta(q)$. 若对某一 $\liminf\limits_{\delta\to 0}$ 能被 $\lim\limits_{\delta\to 0}$ 替代, 则称 $\widetilde{\theta}_R(q)$ 是格正则. 类似地,

$$\widetilde{f}_R(y) = \lim_{\varepsilon\to 0}\limsup_{\delta\to 0}\frac{\log \# K_\delta(y,\varepsilon)}{-\log\delta},$$

其中

$$K_\delta(y,\varepsilon) = \left\{k \in K_\delta : y - \varepsilon < \frac{\log\mu[\overline{B}_\delta(k)]}{\log\delta} \leqslant y + \varepsilon\right\}.$$

同样地, 如果 $\limsup\limits_{\delta\to 0}$ 能被 $\lim\limits_{\delta\to 0}$ 替代, $\widetilde{f}_R(y)$ 被称作格正则. 然后证明了

$$\widetilde{\theta}_R(q) = \inf_y\{qy - \widetilde{f}_R(y)\}, \quad q \neq 0.$$

进一步, 若 $\widetilde{\theta}_R(q)$ 是格正则的, 且为在 $\mathbb{R}$ 中可微的凸函数, 则

$$\widetilde{f_R}(y) = \lim_{\varepsilon\to 0}\lim_{\delta\to 0}\frac{\log \# K_\delta(y,\varepsilon)}{-\log\delta} = \inf_q\{qy - \widetilde{\theta}_R(q)\}.$$

如果 $\widetilde{\theta}_R(q)$ 在 $q \neq 0$ 可微, 且 $y_q = \widetilde{\theta}'_R(q)$, 则

$$\widetilde{\theta}_R(q) = qy_q - \widetilde{f}_R(y_q).$$

2.6.4　上下界

Brawn 等在 1992 年分析了单位区间的格子覆盖. 他们研究了下列集合的 Hausdorff 和 Packing 维数:

$$\{x \in [0,1), \liminf_{n\to\infty} Y_n(x) \leqslant y\}, \quad \{x \in \operatorname{supp}\mu, \liminf_{n\to\infty} Y_n(x) \geqslant y\},$$
$$\{x \in [0,1), \limsup_{n\to\infty} Y_n(x) \leqslant y\}, \quad \{x \in \operatorname{supp}\mu, \limsup_{n\to\infty} Y_n(x) \geqslant y\},$$

其中 $Y_n(x)$ 由等式 (2.10) 给出. □

2.7　点中心情形的结论的复习

本节简要复习点中心结构情形下的一些结论. 一般来讲, 这些结论反映了测度的局部行为的上下极限分别与 Hausdorff 维数和 Packing 维数的关系.

2.7.1　定理 (Young, 1982; Pesin,1993)

设 $S_\delta(x)$ 是一个中心为 x 半径为 δ 的闭球. 如果 μ 是 $\mathcal{B}(\mathcal{X})$ 中的概率测度, 并且 $\lim\limits_{\delta\to 0}\dfrac{\log\mu[S_\delta(x)]}{\log\delta}$ 存在且 μ 几乎处处恒为常数, 则极限为 σ, 且

$$\sigma = \inf\{\dim_H(E) : \mu(E) = 1\}. \tag{2.14}$$

注意到, 通常 E 不是闭集, 所以 μ 的支撑的 Hausdorff 维数可能严格大于 σ.

1986 年及 1991 年, Cutler 和 Dawson 在 1989 年的系列文章中, 给出测度 μ 的局部行为及它们与 Hausdorff 维数和 Packing 维数的关系的更细致的描述. 在定义 2.5.5 中, 在假设 $\lim\limits_{\delta\to 0} Y_\delta(x)$ 存在的前提下给了 $F(y)$ 一个简单定义. Cutler 分别考虑 liminf 和 limsup 的定义, 它们分别与 Hausdorff 维数及 Packing 维数有关, 并且与维数分布有关. 这些想法最早由 Gacs 在 1973 年提出.

2.7.2 定义 (Cutler, 1991)

1) 定义概率测度 μ_H 和 μ_P 分别为

$$\mu_H([0,y]) = \sup\{\mu(E) : \dim_H(E) \leqslant y\},$$
$$\mu_P([0,y]) = \sup\{\mu(E) : \dim_P(E) \leqslant y\}.$$

2) 若 $\mu_H = \mu_P$, 称测度 μ 为维数正则.

3) 如果存在 $y_1 \geqslant 0$ 使 $\mu_H = \delta_{y_1}$(或 $\mu_p = \delta_{y_1}$), 则称测度 μ 有精确 Hausdorff(或 Packing) 维数, 其中 δ_{y_1} 是 y_1 的单位质量.

4) 分别定义局部映射 $Y_{\inf} : \mathcal{X} \to [0,\infty)$ 及 $Y_{\sup} : \mathcal{X} \to [0,\infty)$ 为

$$Y_{\inf} = \liminf_{\delta\to 0} \frac{\log\mu[S_\delta(x)]}{\log\delta},$$
$$Y_{\sup} = \limsup_{\delta\to 0} \frac{\log\mu[S_\delta(x)]}{\log\delta}.$$

维数分布 μ_H 和 μ_P 如上定义. 下列结果指出 $\lim\limits_{\delta\to 0} Y_\delta(x), \mu$-a.s. 存在的充要条件是两种维数的分布是相同的, 其中 $\lim\limits_{\delta\to 0} Y_\delta(x)$ 刻画了 μ 的局部行为. 这样, Hausdorff 维数与 liminf, Packing 维数与 limsup 之间有密切的关系.

2.7.3 定理 (Cutler, 1991)

$Y_{\inf} = Y_{\sup}, \mu$-a.s. 当且仅当 μ 是维数正则的. 进一步, μ 有精确 Hausdorff(或 Packing) 维数 y_1 当且仅当 $Y_{\inf} = y_1, \mu$-a.s. (或 $Y_{\sup} = y_1, \mu$-a.s.).

2.7.4 推论 (Cutler, 1991)

$$\lim_{\delta\to 0} \frac{\log\mu[S_\delta(x)]}{\log\delta} = y_1 = \text{常数},$$

μ-a.s. 当且仅当 μ 是维数正则的且有精确维数 y_1. 在这种情形下 $y_1 = \sigma$, 其中 σ 由等式 (2.14) 给出.

□

2.7.5 定理 (Cutler, 1991, 657 页)

定义集合

$$G_{\inf}(y)=\{x:Y_{\inf}(x)\leqslant y\},$$
$$G_{\sup}(y)=\{x:Y_{\sup}(x)\leqslant y\}.$$

下面的关系成立

1) $\dim_H G_{\inf}(y)\leqslant y$, $\dim_P G_{\sup}(y)\leqslant y$.

2) $\mu_H([0,y])=\mu(G_{\inf}(y)), \mu_P([0,y])=\mu(G_{\sup}(y))$.

此外, 还可以考虑 $G'_{\sup}(y)=\{x:Y_{\sup}(x)\geqslant y\}$, 则 $F(y)=G_{\inf}(y)\cap G'_{\sup}(y)$.

2.7.6 平均局部行为

如果点 x 由 μ 随机确定, 则 $Y_{\inf}(x)$ 和 $Y_{\sup}(x)$ 可看做分别由 μ_H 和 μ_P 确定的随机变量. 分别定义平均 Hausdorff 与 Packing 局部极限为

$$E[Y_{\inf}(x)]=\int Y_{\inf}(x)\mu(dx)=\int_0^\infty y\mu_H(dy),$$
$$E[Y_{\sup}(x)]=\int Y_{\sup}(x)\mu(dx)=\int_0^\infty y\mu_P(dy).$$

$E[Y_{\inf}(x)]=E[Y_{\sup}(x)]$ 当且仅当 μ 是维数正则的. 如果 μ 具有精确维数 σ, 则 $\sigma=E[Y_{\inf}(x)]=E[Y_{\sup}(x)]$. 以上两个定义对于 D_1 有相似的解释, 并且它们在许多情形下相同. 如果测度是维数正则的, 那么唯一的不同是积分和关于 δ 的极限的交换方式. Cutler 在 1991 年定义了 $\mu[S_\delta(x)]$ 的 L^q 形式, 它与 Rényi 维数 D_q 相联系. 类似于格情形下定理 2.3.3 的结果.

2.7.7 定理 (Cutler, 1991, 659 页)

设 $\mathcal{X}\subset\mathbb{R}^d$ 为紧集, μ 为定义在 Borel 集族 $\mathcal{B}(\mathcal{X})$ 上的概率测度. 令 $\theta^-(q)$ 和 $\theta^+(q)$ 为等式 (2.5) 的 $\theta(q)$ 定义中分别用 $\liminf\limits_{\delta\to0}$ 和 $\limsup\limits_{\delta\to0}$ 替换 $\lim\limits_{\delta\to0}$ 得到, 则下面关系成立.

1) 对于 $q\neq1$,

$$\left\{\frac{1}{q-1}\theta^-(q)\right\} \text{ 和 } \left\{\frac{1}{q-1}\theta^+(q)\right\}$$

是关于 q 的减函数.

2) 对所有满足 $r<1<q$ 的 r 和 q,

$$\frac{1}{r-1}\theta^-(r)\geqslant E[Y_{\inf}(x)],\quad E[Y_{\sup}(x)]\geqslant\frac{1}{q-1}\theta^+(q).$$

3) 如果 μ 是维数正则的, 则当 $r < 1 < q$ 时,

$$\frac{1}{r-1}\theta^{-}(r) \geqslant E[Y_{\inf}(x)] = E[Y_{\sup}(x)] \geqslant \frac{1}{q-1}\theta^{+}(q). \qquad \square$$

Cutler 在 1995 年进一步分别讨论了涉及上、下局部映射 $Y_{\sup}(x)$ 和 $Y_{\inf}(x)$ 的 Hausdorff 维数和 Packing 维数的对偶表示. 对“强”、“弱”情形都作了讨论.

Pesin (1993) 和 Olsen (1994,24-26 页) 讨论了多种广义维数的定义以及它们的关系.

第 3 章　多项分布测度

3.1　引　　言

多项分布测度是例 1.2.1 中讨论的 Cantor 测度的一种推广, 这个例子既得到了等式 (2.2) 中的 $\widetilde{\theta}(q)$ 的表达, 也得到了等式 (2.7) 中的重分形谱 $\widetilde{f}(y)$ 的表达. 接着利用 Legendre 变换说明 $\widetilde{\theta}(q)$ 和 $\widetilde{f}(y)$ 确实是相关的. 利用 Lagrange 乘子这样做相对简单些. Legendre 变换也出现在大偏差理论的研究中, 在本书的第二部分将说明利用该理论可以得到 Legendre 变换关系中的充分条件在更一般的情形中也是成立的.

通过将单位区间不断的分割并重新进行质量分配, 可以构造单位区间上的多项分布测度. 首先将单位区间分成 b 个等长的子区间, 这里 $b \geqslant 2$. 每个这样的子区间再分成 b 个子区间, 以此类推. 处在第 n 级的各个子区间可以唯一的由长度为 n 的 b-进制数表达. 进一步, 各子区间的测度大小也可以通过其对应的 b-进制数中的数字组合给出定义. Mandelbrot (1989) 也讨论了这个例子.

3.1.1　测度的构造

记 $\Omega = \{0, 1, \cdots, (b-1)\}$, 其中 $b \in \{2, 3, \cdots\}$ 是给定的. 给每个元素 $\omega \in \Omega$ 赋予一个概率 p_ω, 满足

$$\sum_{\omega=0}^{b-1} p_\omega = 1, \quad p_\omega \geqslant 0.$$

令 $\Omega_0 = \{\omega \in \Omega : p_\omega > 0\}$, $s = \sharp\Omega_0$, Ω^n 和 Ω_0^n 分别是 Ω 和 Ω_0 的 n 阶直积.

下面考虑单位区间 $[0,1]$ 上迭代结构以及概率测度 μ 的构造. 在第一次迭代中, 将 $[0,1]$ 分成 b 个没有重叠的闭子区间. 按比例 $p_0, \cdots, p_{b-1}$ 依次给这 b 个子区间分配质量. 在接下来的每一步迭代过程中, 将每个子区间都分割成 b 个更小的子区间, 并且每个子区间分割后的 b 小块也按比例 $p_0, \cdots, p_{b-1}$ 依次分配质量. 在迭代了 n 次以后, 区间 $[0,1]$ 分割为 b^n 个等长的子区间, 其中有 s^n 个区间具有非零质量.

将序列 $\omega \in \Omega^n$ 看成是区间 $[0,1]$ 上的 b-进制有理数, 则每个序列对应于 b^n 个子区间 (它们的宽度都是 $\delta_n = b^{-n}$) 中的某一个. 也就是说, 在每个确定子区间内部的所有数在作 b 进制展开后前面 n 位都具有相同的表达. 第 n 级中的各个子区

间可以表示成 $J_n(\omega)$, 其中

$$J_n(\omega)=\left[\sum_{j=1}^{n}\frac{\omega_j}{b^j},\ b^{-n}+\sum_{j=1}^{n}\frac{\omega_j}{b^j}\right],$$

这里 $\omega_j\in\Omega$. 可以定义概率测度 μ, 它在每个特定区间上测度为

$$\mu[J_n(\omega)]=p_{\omega_1}\cdots p_{\omega_n}=p_0^{n\Phi_0(\omega)}p_1^{n\Phi_1(\omega)}\cdots p_{b-1}^{n\Phi_{b-1}(\omega)},\tag{3.1}$$

其中 $\Phi_\alpha(\omega)$ 是序列 $\omega\in\Omega^n$ 中等于 α 的那些数字所占的比例. 因而, Ω_0 的 n 阶直积为 $\Omega_0^n=\{\omega\in\Omega^n:\mu[J_n(\omega)]>0\}$.

3.1.2　例子

例 1.2.1 中 Cantor 测度就是一个多项分布测度的例子. 在这个特殊情形中 $b=3;\ p_0=\dfrac{1}{3},p_1=0,p_2=\dfrac{2}{3}$, 因而 $s=2$. □

3.2　局 部 性 态

在例 2.1.1 中, 在一个 Lebesgue 测度为 0 的集合上给出了一个概率测度. 这样的测度没有概率密度函数, 因而无法利用传统的方法对其研究. 刻画这样的测度有个可供选择的方法就是描述下面将要讨论的测度局部性态.

对一个给定的 n, 考虑那些具有非零测度的区间 $J_n(\omega)$, 即 $\omega\in\Omega_0^n$. 对于这些区间, 可以用 $Y_n(\omega)$ 来描述它们的局部性态, 这里

$$\begin{aligned}Y_n(\omega)&=\frac{\log\mu[J_n(\omega)]}{\log\delta_n}\\&=\frac{-1}{n}\log_b\left[\prod_{\alpha\in\Omega_0}p_\alpha^{n\Phi_\alpha(\omega)}\right]\\&=-\sum_{\alpha\in\Omega_0}\Phi_\alpha(\omega)\log_b p_\alpha.\end{aligned}\tag{3.2}$$

我们希望把 Y_n 看成一个离散随机变量. 也就是, 当以同等概率从那些具有正测度的区间中随机的选择一个区间 $J_n(\omega)$ 时, Y_n 等于 y 的概率即 $\Pr\{Y_n=y\}$ 是多少?

注意到离散属性, 可以简化这个问题. 例如令 $b=2,\ n=4$, 则 $\Phi_\alpha(\omega)=0,\dfrac{1}{4},\dfrac{1}{2},\dfrac{3}{4},1$ $(\alpha=0,1)$. 当 $\phi_0,\ \phi_1$ 取好特定的值时, 符合条件的子区间也就确定, 就这些子区间的数目而言上述概率分布的表达得以简化. 例如 $\Phi_0=\Phi_1=\dfrac{1}{2}$ 时, 则长度为 $n=4$ 的 b 进制表达式中有 6 个符合前述赋值条件:

$$\begin{matrix} 0.0011 & 0.0101 & 0.0110 \\ 0.1010 & 0.1100 & 0.1001. \end{matrix}$$

定义 $(b \times 1)$ 向量集 $\Psi(y)$ 如下

$$\Psi(y) = \left\{ (\phi_0, \cdots, \phi_{b-1}) : y = -\sum_{\omega \in \Omega_0} \phi_\omega \log_b p_\omega \text{ 且 } \sum_{\omega \in \Omega_0} \phi_\omega = 1 \right\}, \tag{3.3}$$

定义它的子集 $\Psi_n(y)$ 为

$$\Psi_n(y) = \{(\phi_0, \cdots, \phi_{b-1}) \in \Psi(y) : n\phi_\omega \text{是非负整数}, \forall \omega \in \Omega_0\}.$$

如果 $s = 2$, $p_{\omega_1} \neq p_{\omega_2}$, 这里 $\Omega_0 = \{\omega_1, \omega_2\}$, 则 $\Psi(y)$ 只有唯一的点. 对给定好的 $(\phi_0, \cdots, \phi_{b-1}) \in \Psi_n(y)$, 第 n 级子区间中满足该赋值条件的区间个数 $N_n(\phi_0, \cdots, \phi_{b-1})$ 为

$$N_n(\phi_0, \cdots, \phi_{b-1}) = \frac{n!}{(n\phi_0)! \cdots (n\phi_{b-1})!}.$$

由此可得

$$\Pr\{Y_n = y\} = \begin{cases} 0, & \text{若 } \Psi_n(y) = \varnothing, \\ \dfrac{1}{s^n} \displaystyle\sum_{(\phi_0, \cdots, \phi_{b-1}) \in \Psi_n(y)} N_n(\phi_0, \cdots, \phi_{b-1}), & \text{若 } \Psi_n(y) \neq \varnothing. \end{cases}$$

需要指出: $\Pr\{Y_n = y\} \to 0$, 除非 $y = y_0 = (1/s) \sum_{\omega \in \Omega_0} \log_b p_\omega$. 然而, 可以根据 $\Pr\{Y_n = y\} \to 0$ 的速率的幂律将单位区间分成不同的部分.

3.2.1 重分形谱

重分形谱记作 $\widetilde{f}(y)$, 它描述了 Y_n 概率分布的收敛速度. 处于第 n 级、具有正测度 μ 且满足 $Y_n = y$ 的那些子区间的个数为 $s^n \Pr\{Y_n = y\}$, 这些子区间的宽度均为 $\delta_n = b^{-n}$(见图 1.2). 因而, 前面所讨论的例子中, 重分形谱为

$$\begin{aligned} \widetilde{f}(y) &= \lim_{n \to \infty} \frac{\log(s^n \Pr\{Y_n = y\})}{-\log \delta_n} \\ &= \lim_{n \to \infty} \frac{1}{n} \log_b \sum_{(\phi_0, \cdots, \phi_{b-1}) \in \Psi_n(y)} N_n(\phi_0, \cdots, \phi_{b-1}). \end{aligned} \tag{3.4}$$

注意到

$$\begin{aligned} & \max_{(\phi_0, \cdots, \phi_{b-1}) \in \Psi_n(y)} N_n(\phi_0, \cdots, \phi_{b-1}) \\ \leqslant & \sum_{(\phi_0, \cdots, \phi_{b-1}) \in \Psi_n(y)} N_n(\phi_0, \cdots, \phi_{b-1}) \\ \leqslant & (\sharp \Psi_n(y)) \max_{(\phi_0, \cdots, \phi_{b-1}) \in \Psi_n(y)} N_n(\phi_0, \cdots, \phi_{b-1}), \end{aligned}$$

以及 $\sharp\Psi_n(y) \leqslant (n+1)^b$, 所以

$$\widetilde{f}(y) = \max_{(\phi_0,\cdots,\phi_{b-1})\in\Psi(y)} \lim_{n\to\infty} \frac{1}{n}\log_b N_n(\phi_0,\cdots,\phi_{b-1}).$$

根据 Stirling 逼近原理可得

$$\lim_{n\to\infty}\frac{1}{n}\log_b N_n(\phi_0,\cdots,\phi_{b-1}) = -\sum_{\omega\in\Omega_0}\phi_\omega\log_b\phi_\omega, \tag{3.5}$$

于是

$$\widetilde{f}(y) = \max_{(\phi_0,\cdots,\phi_{b-1})\in\Psi(y)} -\sum_{\omega\in\Omega_0}\phi_\omega\log_b\phi_\omega. \tag{3.6}$$

□

等式 (3.4) 中函数 $\widetilde{f}(y)$ 的形式类似于计盒维数, 因为它是统计覆盖盒子的个数, 这些盒子按照事先给定的盒子宽度进行分割, 并且这个宽度趋向于 0. 但是, 它又是不同于计盒维数的, 因为所要覆盖的集合的类型在 n 趋向于无穷大的时候也发生了变化.

3.3　全局平均和 Legendre 变换

描述测度 μ 的一个可行的方法是利用 μ 的 q 次幂, 或者 q 次矩的全局平均. 首先给出定义, 然后说明经过一个 Legendre 变换后这个全局平均和局部性态是相关的.

3.3.1　全局平均

在这个例子中, 等式 (2.2) 定义的关联指数可以表示为

$$\widetilde{\theta}(q) = \lim_{n\to\infty}\frac{\log\sum_{\omega\in\Omega_0^n}\mu^q[J_n(\omega)]}{\log\delta_n}, \tag{3.7}$$

其中 $q\in\mathbb{R}$, $\delta_n = b^{-n}$ 是子区间的宽度. 于是有

$$\begin{aligned}\widetilde{\theta}(q) &= \lim_{n\to\infty}\frac{-1}{n}\log_b\left[\sum_{\omega_1\in\Omega_0}\cdots\sum_{\omega_n\in\Omega_0}p_{\omega_1}^q\cdots p_{\omega_n}^q\right]\\ &= -\log_b\sum_{\omega\in\Omega_0}p_\omega^q.\end{aligned} \tag{3.8}$$

□

在有些文献中, $\widetilde{\theta}(q)$ 经常写成 $\tau(q)$. 之所以用 θ 是因为在本书的第三部分, 考虑了 $\theta(q)$ 的估计方法. 用 θ 替代 τ 便于区分实数 θ, 抽样估计 $\widehat{\theta}$ 以及估计量 Θ.

3.3.2 例子

考虑 $b=10$ 的多项分布测度情形. 如果 b 取小的数值, 譬如 2 或 3, 则一个相对大数值的 p_ω 必然意味着其他 p_ω 取较小的值 (不具一般性). 考虑三种情形：第一种是 $p_\omega=0.1,\ \omega=0,\cdots,9$. 这种情形就是简单的一致分布, $\widetilde{\theta}(q)$ 对应的图像是图 3.1 中斜率为 1 的实线. 第二种情形是 p_ω 中一个取值 0.001, 其他皆取值 0.111. 这个接近于一致分布, 但是它有一些小区域具有非常小但不为零的测度. $\widetilde{\theta}(q)$ 对应的图像在 $q<0$ 时远离一致测度, 在 $q>0$ 时与一致测度类似. 最后一种是 p_ω 中一个取值 0.37, 其他皆取值 0.07. 这一情形里 $\widetilde{\theta}(q)$ 对应的图像在 $q>0$ 时远离一致测度, 在 $q<0$ 时与一致测度颇为接近.

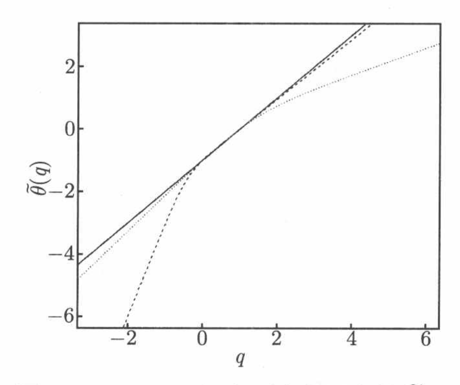


图 3.1 $b=10$ 时, 多项分布测度的 $\widetilde{\theta}(q)$

实线对应 p_ω 皆取值 0.01 的情形; 虚线对应 p_ω 中一个取值 0.001, 其他皆取值 0.111 的情形; 点划线对应 p_ω 中一个取值 0.37, 其他皆取值 0.07 的情形

当所有不为零的 p_ω 均相等时, 测度 μ 在其支撑上是一致分布的. 此时, 根据等式 (3.8) 可知, $\widetilde{\theta}(q)$ 将是以 $\widetilde{D}_0$ 为斜率的直线. 当测度在其支撑上不是一致分布的, $\widetilde{\theta}(q)$ 成为向下凹的曲线, 即以 $q=0$ 为支点向下拉线条的两端, 但是曲线仍然是随着 q 的增加而单调递增的. 在 $q>0$ 且逐渐增加时, 如果某些区域具有相比其他区域大得多的测度, 则线条将向下拉远离原先在一致情形中所处的位置. 反之, 在 $q<0$ 且逐渐减小时, 如果某些区域具有非常小的非零测度, 则线条将下拉远离原先在一致情形中所处的位置. 总之, 函数 $\widetilde{\theta}(q)$ 可以解释成是某个测度偏离在其支撑上一致分布测度的程度. 当 $q>0$ 时, 它描述的是具有较大测度的大量区域; 当 $q<0$ 时, 它描述了测度较小但不为零的区域. 这一行为与定理 2.3.3 中所表达的想法相关. □

3.3.3 Legendre 变换

现在说明 $\widetilde{f}(y)$ 和 $\widetilde{\theta}(q)$ 在 Legendre 变换形式下的关系. 从等式 (3.6) 所表示的

$\widetilde{f}(y)$ 开始, 考虑下式在约束条件下的最大值

$$-\sum_{\omega\in\Omega_0}\phi_\omega\log_b\phi_\omega,$$

其中约束条件为 $(\phi_0,\cdots,\phi_{b-1})\in\Psi(y)$, 即

$$\sum_{\omega\in\omega_0}\phi_\omega=1 \tag{3.9}$$

以及

$$y=-\sum_{\omega\in\Omega_0}\phi_\omega\log_b p_\omega. \tag{3.10}$$

引入两个 Lagrange 乘子 λ 和 q, 考虑函数

$$Q=-\sum_{\omega\in\Omega_0}\phi_\omega\log_b\phi_\omega+q\left(y+\sum_{\omega\in\Omega_0}\phi_\omega\log_b p_\omega\right)+\lambda\left(-1+\sum_{\omega\in\Omega_0}\phi_\omega\right).$$

于是

$$\frac{\partial Q}{\partial\phi_\omega}=-\log_b\phi_\omega-1+q\log_b p_\omega+\lambda,\quad \omega\in\Omega_0, \tag{3.11}$$

利用两个约束条件和 $\partial Q/\partial\phi_\omega=0$, 解出 ϕ_ω, $\omega\in\Omega_0$. 有 $s+2$ 个方程, $s+3$ 个参数, 即 s 个 ϕ_ω, λ, q 和 y, 其中 $s=\sharp\Omega_0$. 把 y 表示成 q 的函数 y_q. 令式 (3.11) 等于 0 可得

$$\log_b\phi_\omega=\lambda-1+q\log_b p_\omega,$$

也就是, $\phi_\omega=b^{\lambda-1}p_\omega^q$. 由式 (3.9) 得

$$\sum_{\omega\in\Omega_0}\phi_\omega=\sum_{\omega\in\Omega_0}b^{\lambda-1}p_\omega^q=1.$$

所以对给定的 y 和 q, 最大值在如下情形取到

$$\phi_\omega=\frac{p_\omega^q}{\sum\limits_{i\in\Omega_0}p_i^q},\quad \omega\in\Omega_0. \tag{3.12}$$

结合式 (3.10) 和 (3.12) 得

$$y_q=\frac{-1}{\left(\sum\limits_{i\in\Omega_0}p_i^q\right)}\sum_{\omega\in\Omega_0}p_\omega^q\log_b p_\omega.$$

再结合式 (3.6) 有

$$\widetilde{f}(y_q)=-\sum_{\omega\in\Omega_0}\frac{p_\omega^q}{\left(\sum\limits_{i\in\Omega_0}p_i^q\right)}\log_b\left(\frac{p_\omega^q}{\sum\limits_{j\in\Omega_0}p_j^q}\right)$$

$$= \frac{-q}{\sum\limits_{i\in\Omega_0} p_i^q} \sum_{\omega\in\Omega_0} p_\omega^q \log_b p_\omega + \log_b \sum_{j\in\Omega_0} p_j^q$$

$$= \frac{-q}{\sum\limits_{i\in\Omega_0} p_i^q} \sum_{\omega\in\Omega_0} p_\omega^q \log_b p_\omega - \widetilde{\theta}(q). \tag{3.13}$$

因而

$$\widetilde{\theta}(q) = qy_q - \widetilde{f}(y_q) = -\log_b \sum_{\omega\in\Omega_0} p_\omega^q. \tag{3.14}$$

这是由 $\widetilde{\theta}(q) = \inf_y\{qy - \widetilde{f}(y)\}$ 给出的 Legendre 变换的一个解. 函数 y_q 表示下确界取到 (即 $\widetilde{f}(y)$ 达到最大) 时, y 作为 q 的函数的取值.

另外, 可以看到

$$y_{\min} = \lim_{q\to\infty} y_q = \min_{\omega\in\Omega_0}(-\log_b p_\omega), \tag{3.15}$$

以及

$$y_{\max} = \lim_{q\to-\infty} y_q = \max_{\omega\in\Omega_0}(-\log_b p_\omega). \tag{3.16}$$

3.3.4 例子——Cantor 测度

取 $b=3$ 以及 $p_0 = \frac{1}{3}$, $p_1 = 0$, $p_2 = \frac{2}{3}$. 则 $\widetilde{f}(y)$ 的 Legendre 变换和函数 y_q 的图像分别如图 3.2 和图 3.3 所示. 注意到下确界在函数 $\widetilde{f}(y)$ 和直线 qy 有相同斜率 (斜率 $= q$) 时取到. 逆 Legendre 变换的图像如图 3.4 所示. □

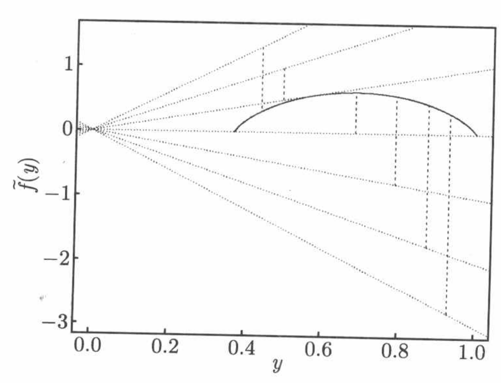


图 3.2 Cantor 测度: $\widetilde{f}(y)$ 的 Legendre 变换

$p_0 = \frac{1}{3}$, $p_1 = 0$, $p_2 = \frac{2}{3}$ 时多项分布测度的 $\widetilde{f}(y)$(实线). 点划线的斜率为 $q = -3, -2, \cdots, 2, 3$. 点划线到曲线 $\widetilde{f}(y)$ 的最小距离在 $y = y_q$(见图 3.3) 处取到, 此处函数 $\widetilde{f}(y)$ 的导数等于 q. 图中垂直的虚线表示在 $y = y_q$ 处达到的那些最小距离, 它们的长度为 $\widetilde{\theta}(q) = \inf_y\{qy - \widetilde{f}(y)\}$

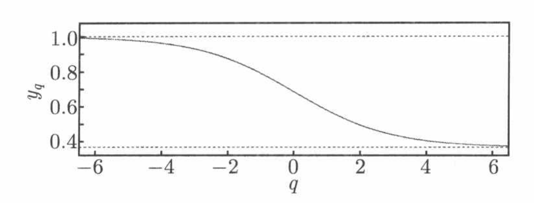


图 3.3　Cantor 测度：函数 y_q

$p_0 = \frac{1}{3}$, $p_1 = 0$, $p_2 = \frac{2}{3}$ 时多项分布测度的 y_q 函数. 函数 $\widetilde{f}(y)$(见图 3.2) 在 $y = y_q$ 的导数等于 q. 两条渐近线为 $y_{\min} = \lim_{q\to\infty} y_q$ 和 $y_{\max} = \lim_{q\to-\infty} y_q$, 在 y 取这些值时, 函数 $\widetilde{f}(y)$ 与水平轴相交

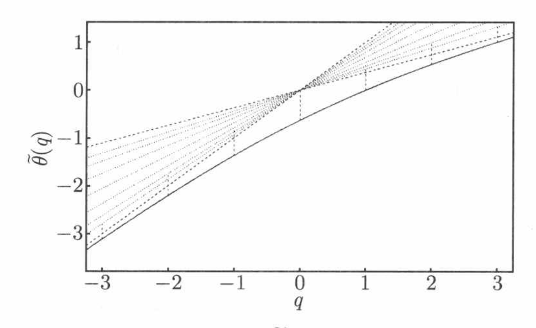


图 3.4　Cantor 测度：$\widetilde{\theta}(q)$ 的 Legendre 变换

$\widetilde{\theta}(q)$ 如实线所示. 当 q 增大且为正数时, $\widetilde{\theta}(q)$ 趋近于直线 $qy_{\min}$, 当 q 减小且为负数时, $\widetilde{\theta}(q)$ 趋近于直线 $qy_{\max}$(两条对角的虚线). 点划线分别表示斜率为 $y = y_q$, $q = -3, -2, \cdots, 2, 3$ 的直线 (见图 3.3). $\widetilde{\theta}(q)$ 的斜率为 y_q. 曲线 $\widetilde{\theta}(q)$ 与直线 qy 最小的距离 $\widetilde{f}(y)$ (图中垂直的虚线所示) 将在满足直线斜率为 y_q(即 $y_q = y$) 的 q 处达到, 即 $\widetilde{f}(y_q) = qy_q - \widetilde{\theta}(q)$ 是 $\widetilde{f}(y) = \inf_y\{qy - \widetilde{\theta}(q)\}$ 的解

3.4　分形维数

定义映射 $g_n : [0,1] \to \Omega^n$, 这里 $g_n(x)$ 表示 x 的 b-进制展开式的前 n 位数字. 想要估计如下集合的盒维数或者 Hausdorff 维数

$$\widetilde{F}(y) = \left\{x \in [0,1] : \lim_{n\to\infty} Y_n(g_n(x)) = y\right\}.$$

3.4.1　Eggleston 定理 (Eggleston, 1949)

设 $\Phi_\alpha(x)$ 表示 x 的 b-进制展开式中等于 α 的那些数字所占的渐近比例, 令

$$M(\phi_0, \cdots, \phi_{b-1}) = \{x \in [0,1] : \Phi_0(x) = \phi_0, \cdots, \Phi_{b-1}(x) = \phi_{b-1}\},$$

其中 $\sum_{\omega\in\Omega} \phi_\omega = 1$. 则 $M(\phi_0, \cdots, \phi_{b-1})$ 的 Hausdorff 维数为

$$\dim_H M(\phi_0, \cdots, \phi_{b-1}) = -\sum_{\omega\in\Omega_0} \phi_\omega \log_b \phi_\omega. \tag{3.17}$$

也可参见 Falconer (1990, §10.1). □

已经知道几乎所有的数 (在 Lebesgue 测度意义下)在任意 b 进制表达下都是正规的, 即集合 $M\left(\frac{1}{b},\cdots,\frac{1}{b}\right)$ 的 Lebesgue 测度为 1, 从而维数也为 1. 这与公式 (3.17) 是一致的. 如果 $\phi_\omega>0$, $\forall\omega\in\Omega$, 则集合 $M(\phi_0,\cdots,\phi_{b-1})$ 在 $[0,1]$ 中稠密, 因而 $\dim_B M(\phi_0,\cdots,\phi_{b-1})=1$. 不幸的是, 这是计盒维数的一个缺憾 (见性质 A.3.4 和例 A.3.5).

3.4.2 例子

考虑特殊情形 $s=\sharp\Omega_0=2$, 则 $\exists i,j:i\neq j$ 满足 $p_i=1-p_j$, 同时当 $m\neq i$ 且 $m\neq j$ 时 $p_m=0$. 由式 (3.2) 得

$$Y_n(\omega)=\begin{cases}-\Phi_i(\omega)\log_b\left(\dfrac{p_i}{p_j}\right)-\log_b p_j, & \text{若 } p_i\neq p_j,\\ -\log_b p_j, & \text{若 } p_i=p_j.\end{cases}$$

此时容易看出, 给定 p_i, p_j 且 $p_i\neq p_j$ 时, $\Phi_i(\omega)$ 取值唯一, 因而 $\Phi_j(\omega)$ 也是如此, $\widetilde{F}(y)$ 可以通过这些确定的值表示, 即

$$\widetilde{F}(y)=\{x\in[0,1]:\Phi_i(x)=\phi_i,\ \Phi_j(x)=\phi_j,\ \Phi_m(x)=0,\ \forall m\neq i,j\}.$$

注意根据式 (3.10), y 是与 ϕ_i 和 ϕ_j 相关的, 即 $y=-\phi_i\log_b p_i-\phi_j\log_b p_j$. 式 (3.3) 给出的集合 $\Psi(y)$ 在 $s=2$ 且 $p_i\neq p_j$ 时只含有一个元素, 于是由式 (3.6) 可得

$$\widetilde{f}(y)=-\sum_{\omega\in\Omega_0}\phi_\omega\log_b\phi_\omega=\begin{cases}-\phi_i\log_b\phi_i-\phi_j\log_b\phi_j, & \text{若 } p_i\neq p_j,\\ -\log_b p_j, & \text{若 } p_i=p_j.\end{cases}$$

再结合式 (3.17) 知, $\dim_H\widetilde{F}(y)=\widetilde{f}(y)$. □

3.5 点中心构造

本章刻画多项分布测度的重分形特性的方法是以格子 (固定宽度) 覆盖体系为基础的. 这些结论与从点中心覆盖体系里得到的结论有什么关系? 本节简要讨论了这样的问题.

3.5.1 例子

考虑例 1.2.1 中的 Cantor 测度, $b=3$, $p_0=1-p_2$ 以及 $p_1=0$. 这里让 p_0 取 0 到 1 之间的任意值. 该测度支撑在 Cantor 集上, 这是一个典型的三分 Cantor 集,

中间有三分之一的间隔. 在每次分割步骤中, 分别分配给左边和右边的小区间各 p_0 和 p_2 的质量. 设 μ_n 为这一过程中第 n 步得到的 $\mathcal{K}_n$ 上的测度分布 (见式 (3.1)).

设 X 是一个服从分布 μ 的随机变量, 期望 $E[\mu^{q-1}[S_\delta(X)]]$ 可以通过先估算 $E[\mu_n^{q-1}[S_\delta(X)]]$ 来计算. 每个测度不为零的子区间都具有一致密度 (如例 2.4.4), 故可以将这个积分 (记为 $A_n(q,\delta)$) 表示成这些小区间上各自积分的求和, 即

$$A_n(q,\delta)=\sum_{\omega\in\Omega_0^n}\int_{x\in J_n(\omega)}\mu_n^{q-1}[S_\delta(x)]\mu_n(dx),$$

其中 δ 取得充分小 (即要求 $\delta<3^{-n}$), Ω_0^n 是数字 0 和 2 组成的长度为 n 的排列的全体, 而 $J_n(\omega)$ 是与 b 进制序列 $\omega\in\Omega_0^n$ 对应的子区间. 然而,

$$D_q\neq\lim_{\delta\to 0}\frac{\log\{\lim_{n\to\infty}A_n(q,\delta)\}}{\log\delta}.$$

这是因为当取第一个极限 $n\to\infty$ 时, 间隔也越来越小. 如果 δ 是固定的, 就会出现半径为 δ 的覆盖小球完全盖过这个间隔, 另一方面又覆盖了支撑集的一部分. 交换两个极限的次序也不能使之成为等式, 这是因为如果可以改成等号, 固定 n, 则 μ_n 由一些本质上类型都是一致的测度所组成, 这些测度被若干间隔所分开. 就像在例 2.4.4 中那样, 在这个测度上对 δ 取极限容易得到对任何 q, $D_q=1$, 这显然不可能.

下面仅考虑 $q=2$ 的情形. 为了找到 D_2 的一个解析解, 需要回到 §2.4 开始部分的想法, 特别是式 (2.4). 设 X_1 和 X_2 是 $\mathcal{K}_n$ 上服从分布 μ_n 的独立的随机变量, 则由式 (2.4) 可得

$$\int\mu_n[S_\delta(x)]\mu_n(dx)=\Pr\{|X_1-X_2|\leqslant\delta\}.$$

令 $h_n(y)$ 为 X_1-X_2 的密度, $g_n(y)$ 为 $|X_1-X_2|$ 的密度. 设 $\mathcal{K}_0$ 的分布是一致分布, 故

$$\begin{aligned}h_0(y)&=1-|y|,\quad -1\leqslant y\leqslant 1,\\ g_0(y)&=2-2y,\quad 0\leqslant y\leqslant 1.\end{aligned}$$

对每个 n, $h_n(y)$ 是由 $h_{n-1}(y)$ 按照某个比例压缩所得, 首先 $h_{n-1}(y)$ 被压缩在区间 $\left[-\frac{1}{3},\frac{1}{3}\right]$ 上, 两边的区间 $\left[-1,-\frac{1}{3}\right]$ 和 $\left[\frac{1}{3},1\right]$ 上也有这个压缩形的复制. 中间一块区间上的压缩比例是 $p_0^2+p_2^2$, 两边区间上的的压缩比例为 p_0p_2, 所以 $h_n(y)$ 满足递推关系

$$h_n(y)=\begin{cases}3(p_0^2+p_2^2)h_{n-1}(3y), & 若\ 0\leqslant|y|\leqslant\dfrac{1}{3},\\ 3p_0p_2h_{n-1}(3|y|-2), & 若\ \dfrac{1}{3}\leqslant|y|\leqslant 1.\end{cases}$$

因此,

$$g_n(y)=\begin{cases}3(p_0^2+p_2^2)g_{n-1}(3y), & 若\ 0\leqslant y\leqslant\dfrac{1}{3},\\ 3p_0p_2g_{n-1}(2-3y), & 若\ \dfrac{1}{3}\leqslant y\leqslant\dfrac{2}{3},\\ 3p_0p_2g_{n-1}(3y-2), & 若\ \dfrac{2}{3}\leqslant y\leqslant 1,\end{cases}$$

于是绝对值点距的概率分布满足如下递推关系:

$$G_n(y)=\begin{cases}(p_0^2+p_2^2)G_{n-1}(3y), & 若\ 0\leqslant y\leqslant\dfrac{1}{3},\\ 1-p_0p_2-p_0p_2G_{n-1}(2-3y), & 若\ \dfrac{1}{3}\leqslant y\leqslant\dfrac{2}{3},\\ 1-p_0p_2+p_0p_2G_{n-1}(3y-2), & 若\ \dfrac{2}{3}\leqslant y\leqslant 1.\end{cases} \tag{3.18}$$

故当 $0<y<3^{-n}$ 时, $G_n(y)=(p_0^2+p_2^2)^nG_0(3^ny)$, 其中 $G_0(y)=2y-y^2$. 令 $G_\infty(y)$ 是这个极限概率分布, 则

$$D_2=\lim_{y\to 0}\frac{\Pr\{|X_1-X_2|\leqslant y\}}{\log y}=\lim_{y\to 0}\frac{G_\infty(y)}{\log y}.$$

根据构造, $G_\infty(3^{-n})=(p_0^2+p_2^2)^nG_0(1)=(p_0^2+p_2^2)^n$, 从而对所有满足 $3^{-(n+1)}\leqslant y\leqslant 3^{-n}$ 的 y 有

$$(p_0^2+p_2^2)^{n+1}\leqslant G_\infty(y)\leqslant(p_0^2+p_2^2)^n,$$

所以

$$\frac{\log[(p_0^2+p_2^2)^{n+1}]}{\log(3^{-n})}\geqslant\frac{\log G_\infty(y)}{\log y}\geqslant\frac{\log[p_0^2+p_2^2)^n]}{\log(3^{-(n+1)})}.$$

对 n 取极限可得

$$D_2=\lim_{y\to 0}\frac{\log G_\infty(y)}{\log y}=-\log_3(p_0^2+p_2^2)=\widetilde{D}_2. \qquad \square$$

事实上, 不难看出上述讨论对任意具有间隔的多项分布测度和所有的 q 阶点距都是成立的, 即 $D_q=\widetilde{D}_q$, $q=2,3,\cdots$.



第二部分

大偏差下的重分形公式

第 4 章 基于格点的重分形

4.1 引言

本章将给出重分形格公式成立的一个充分条件. 在该条件下, 弱和强的意义下(定义 2.5.4) 的重分形公式都成立. 本节先给出整个讨论的一个概要.

设 $\mathcal{X}$ 是 $\mathbb{R}^d$ 的一个子集, $\mathcal{B}(\mathcal{X})$ 是 $\mathcal{X}$ 的 Borel 子集构成的 σ-代数. 考虑概率空间 $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu)$, 这里假设测度 μ 为无原子的, 且可以集中在 $\mathcal{X}$ 的一个维数低于 d 的子集上. 在本章中考虑情形: $\mathcal{X}$ 被逐级的维数为 d、宽度为 δ_n、左闭右开的盒子覆盖. 第 k 个盒子记为 $B_{\delta_n}(k)$, 其中 $k \in K_n$, $K_n = \{k : \mu[B_{\delta_n(k)}] > 0\}$. 对某个序列 $\{\delta_n\}$, 其中当 $n \to \infty$ 时 $\delta_n \to 0$, 来估计逐级的格覆盖. 利用这种格系统进行估计, 式 (2.10) 中由 Y_n 表示的测度 μ 的局部行为就是一个给定盒子内的测度的对数与该盒子宽度的对数比值.

注意第 3 章多项分布测度的例子中, 覆盖是由闭区间, 即 b-进制的闭区间组成. 这里不一样的原因将在例 4.5.11 中讨论.

为了引入大偏差理论, 最好定义辅助概率空间 $(\mathcal{X}, \mathcal{F}_n, \nu^{(n)})$, 这里 $\mathcal{F}_n$ 是由第 n 级的格生成的子 σ-代数, $\nu^{(n)}$ 是任意一个描述盒子取样方式的概率测度. 于是把 Y_n 看成一个随机变量, 其中 $Y_n : \mathcal{X} \to \mathbb{R}$.

包含 Legendre 变换关系的弱重分形公式可以通过大偏差中的 Gärtner-Ellis 定理导出. 在这一情形中, Y_n 收敛于一个退化点, 这个点依赖于样本测度 $\nu^{(n)}$ 的具体形式. 这个收敛类似于弱大数定律, 但是以指数级收敛. 我们会发现重分形谱 $\widetilde{f}(y)$ 与刻画收敛速度的熵函数相关, 而 $\widetilde{\theta}(q)$ 与 Y_n 的一个重整累积生成函数有关. §4.2 建立了研究一个任意样本测度的数学框架. §4.3 讨论了一致样本测度的情形, §4.4 定义了一族样本测度, 使得族里单个成员的重整累积生成函数和熵函数可以通过简单移位型变换联系起来.

在 §4.5 中, 给出使 $\widetilde{f}(y)$ 等于 μ 的局部性态的水平集的 Hausdorff 维数的充分条件. 为此, 要求样本测度 $\nu^{(n)}$ 可以延拓为 $\nu^{(\infty)}$, 这里 $\nu^{(\infty)}$ 是 $\nu^{(n)}$ 的递归极限. 也就是说, 考虑 Y_n 在概率空间 $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \nu^{(\infty)})$ 上的行为. 在该情形下, Y_n 几乎必然收敛于退化点. 这也使得刻画测度 $\nu^{(\infty)}$ 的局部性态要比 μ 简单. 因为有几乎必然收敛, 所以刻画 μ 局部性态的集合与刻画 $\nu^{(\infty)}$ 的局部性态的某些集合在除去一个零测集意义下是一样的. 于是问题本质上简化为确定 $\nu^{(\infty)}$ 的局部性态的水平集的

Hausdorff 维数. 这个可以借助于 Billingsley (1965) 的一个定理, 然后再回到测度 μ 的局部性态.

4.2 大偏差公式

由第 3 章的多项分布测度的例子看到, 可以通过描述测度的局部性态 Y_n 来刻画这个测度, 见公式 3.2. 把 Y_n 看成一个随机变量, 则可以证明它具有一个退化的分布, 然而我们仍然可以通过研究 Y_n 的收敛速率来得到一些信息. 这就是大偏差理论所研究的内容. 附录 B 中给出了大偏差理论的一个介绍, 以及本章中需要用到的一些结论. 在本节先给出 Y_n 的一个一般性的概率论理论框架, 接下来的各小节在此基本框架中进一步讨论. 从辅助概率空间 $(\mathcal{X}, \mathcal{F}_n, \nu^{(n)})$ 开始, 这里 $\nu^{(n)}$ 是某个任意的样本测度.

为了刻画 μ 的性态, 定义一列可测映射 $\{U_n\}$, $n = 1, 2, \cdots$, 其中

$$U_n : (\mathcal{X}, \mathcal{F}_n) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), \tag{4.1}$$

定义为

$$U_n(x) = \begin{cases} -\log \mu[B_{\delta_n}(k(x))], & 若\ k(x) \in K_n, \\ 0, & 若\ k(x) \notin K_n. \end{cases}$$

这里 $k(x)$ 是指包含 x 盒子的指标 k. 同时, 函数 $\{Y_n\}$, $n = 1, 2, \cdots$ 定义为

$$Y_n(x) = \frac{U_n(x)}{-\log \delta_n}.$$

注意, 虽然 Y_n 是定义在 $x \in \mathcal{X}$ 上, 但 Y_n 对同一个盒子中的所有 x 取值是一样的. 我们想通过研究下面集合的大小来描述测度 μ 的性态

$$\{B_{\delta_n} : -\varepsilon < Y_n - y \leqslant \varepsilon\} = \bigcup_{k \in K_n(y,\varepsilon)} B_{\delta_n}(k).$$

见定义 2.5.1, 其大小就是它的 $\nu^{(n)}$ 测度值: 即 $\nu^{(n)}\{B_{\delta_n} : -\varepsilon < Y_n - y \leqslant \varepsilon\}$. 不同的样本测度将决定相应不同的特征.

利用大偏差理论描述映射 $\{Y_n\}$ 主要包含以下几步. 首先是定义累积生成函数, 用 $C(q)$ 表示. 这本质上是 $\widetilde{\theta}(q)$ 的一个一般化形式. 大偏差理论是基于凸集的讨论, 所以为了进一步研究, 需要假设 $C(q)$ 满足一些条件, 这将在广义假设里指出. 接着定义熵函数 $I(y)$, 也就是 $C(q)$ 的 Legendre 变换, 它也是 $\widetilde{f}(y)$ 的一个一般化形式. 如果充分条件满足, 则 Y_n 收敛到的退化点可以方便的用 $C(q)$ 计算出来. 事实上, 它以指数速率收敛. 熵函数 $I(y)$ 描述了收敛速率, 这可由大偏差的界刻画出来.

4.2.1 重整累积生成函数

根据公式 B.5, 若下列极限 $(n \to \infty)$ 存在, 则重整累积生成函数是 $C(q)$, 其中

$$C(q) = \lim_{n\to\infty} \frac{1}{-\log \delta_n} \log E_n[e^{qU_n}],$$

E_n 是指对一个函数取在样本测度 $\nu^{(n)}$ 下的期望. 于是, 对 $q \in \mathbb{R}$,

$$C(q) = \lim_{n\to\infty} \frac{1}{-\log \delta_n} \log \left(\sum_{k\in K_n} \mu^{-q}[B_{\delta_n}(k)]\nu^{(n)}[B_{\delta_n}(k)] \right). \tag{4.2}$$

注意, 这里允许极限取到 $+\infty$. 另外, 令

$$I(y) = \sup_{q\in\mathbb{R}}\{qy - C(q)\}, \quad y > 0. \tag{4.3}$$

$I(y)$ 可以取值 $+\infty$, 它是 $C(q)$ 的 Legendre 变换. □

可以看出 $C(q)$ 本质上类似于定义 2.3.1 中定义的 $\widetilde{\theta}(q)$ 全局平均. 事实上 $\widetilde{\theta}(q)$ 是一个特殊情形: 样本测度给予每个盒子同样权重. 将在 §4.3 考虑这一情形.

4.2.2 广义假设

假设存在一个序列 δ_n, 以及 $\delta_n \to 0$ $(n \to \infty)$, 满足

1) $C(q)$ 满足假设 B.3.8,

2) $C(q)$ 在它的定义域的内部是可微的, $\mathrm{int}\mathcal{D}(C) = \mathrm{int}\{q \in \mathbb{R} : C(q) < \infty\}$, 并且 $C(q)$ 是陡峭的 (定义 B.2.1). □

本章中的讨论要求 $C(q)$ 满足上述广义假设. 对于任意的序列 δ_n, $C(q)$ 一般不满足上述条件. 在第 3 章的多项分布测度情形, 是取 $\delta_n = b^{-n}$.

给定一个充分好的 $C(q)$, 满足上面的广义假设 4.2.2, 则可以说明 Y_n 以指数速率收敛于一个点 y_0, 这个点可以通过重整累积生成函数推导出来. 记 Y_n 为一个部分平均, 即一个部分和除以 n. 那么指数收敛 (定义 B.3.13) 恰好就是弱大数定律, 不过这里是以指数速率收敛. 一个更常见的情形就是把这种想法运用到若干独立正规的随机变量之和, 见例 B.2.4. 这种与弱大数定律间的相似性将在 §4.3 中进一步描述.

4.2.3 定理——指数收敛

如果 $C(q)$ 满足广义假设 4.2.2, 则 $I(y)$ 在 $y_0 = C'(0)$ 处取得唯一的极小值 $I(y_0) = 0$. 进一步相对于测度 $\nu^{(n)}$, $Y_n \xrightarrow{\exp} y_0$(见定义 B.3.13), 并且

$$C(q) = \sup_{y}\{qy - I(y)\}. \tag{4.4}$$

证明　由假设 B.3.8 以及定理 B.3.10 可得 $\inf_y I(y) = 0$. 因为 $C(q)$ 在 $\mathrm{int}\mathcal{D}(C)$ 是可微的 (广义假设 4.2.2), 注意到 $\mathrm{int}\mathcal{D}(C)$ 非空, 且包含 $q = 0$(假设 B.3.8), 所以根据定理 B.3.14 可得 $I(y)$ 具有唯一极小值, 且在 $y_0 = C'(0)$ 处取到, 以及 $Y_n \xrightarrow{\exp} y_0$. 式 (4.4) 由定理 B.3.17 即得. □

因为, 当 $y \neq y_0$ 时 $I(y) > 0$, $I(y)$ 是幂律的指数 (即 $\delta_n^{I(y)}$), 对任意 $y \neq y_0$, $\nu^{(n)}\{B_{\delta_n} : -\varepsilon < Y_n - y \leqslant \varepsilon\}$ 以该幂律为速率趋向于零. 在下面的定理中正式陈述这个结论.

4.2.4　定理——大偏差的界

如果 $C(q)$ 满足广义假设 4.2.2, 则

$$\lim_{n\to\infty} \frac{\log \nu^{(n)}\{B_{\delta_n} : -\varepsilon < Y_n - y \leqslant \varepsilon\}}{-\log \delta_n} = -\inf_{z\in(y-\varepsilon, y+\varepsilon]} I(z),$$

其中 $\varepsilon > 0$, $(y-\varepsilon, y+\varepsilon] \subseteq \mathcal{D}(I)$. 进一步, 对 ε 取极限得

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} \frac{\log \nu^{(n)}\{B_{\delta_n} : -\varepsilon < Y_n - y \leqslant \varepsilon\}}{-\log \delta_n} = -I(y), \quad y \in \mathrm{int}\mathcal{D}(I). \tag{4.5}$$

证明　由广义假设 4.2.2 中的 (1) 以及 Gärtner-Ellis 定理 B.3.10 可得大偏差的上界成立. 类似地, 由广义假设 4.2.2 中的 (2) 以及 Gärtner-Ellis 定理 B.3.10 可得大偏差的下界成立, 并且 $I(y)$ 是凸且闭的. 于是由引理 B.3.12 可得 $I(y)$ 在 $\mathcal{D}(I)$ 上是连续的. 第一个等式由定理 B.3.6 可得; 对 ε 取极限可得等式 (4.5). □

4.3　均匀空间样本测度

这一节将把 §4.2 的结论运用到样本测度是均匀的情形. 这里均匀的意思是指在逆极限过程中 ($\nu^{(n)}$ 中的 n 有限), 样本测度赋予每个具有非零 μ 测度的盒子同样的权重. 对这种情形用 $\nu_0^{(n)}$ 表示样本测度 $\nu^{(n)}$, 有

$$\nu_0^{(n)}[B_{\delta_n(k)}] = \begin{cases} \dfrac{1}{\sharp K_n}, & 若\ k \in K_n, \\ 0, & 若\ k \notin K_n. \end{cases}$$

类似地, $C(q)$, $I(y)$ 分别用 $C_0(q)$, $I_0(y)$ 表示. 这是本书中普遍考虑的情形.

4.3.1　定理——重整累积生成函数

如果 $\widetilde{\theta}(0)$ 存在, 也就是 $-1 \in \mathcal{D}(C_0)$, 则

$$C_0(q) = \widetilde{\theta}(0) - \widetilde{\theta}(-q), \tag{4.6}$$

这里 $\widetilde{\theta}(q)$ 由定义 2.3.1 给出.

证明 根据等式 (4.2),

$$C_0(q) = \lim_{n\to\infty} \frac{1}{-\log \delta_n} \log\left(\frac{1}{\sharp K_n}\sum_{k\in K_n}\mu^{-q}[B_{\delta_n}(k)]\right),$$

因为 $C_0(-1) = \widetilde{\theta}(0)$, 则 $-1 \in \mathcal{D}(C_0)$ 当且仅当 $\widetilde{\theta}(0)$ 存在且有限, 所以如果 $-1 \in \mathcal{D}(C_0)$, 则

$$C_0(q) = \lim_{n\to\infty}\frac{\log \sharp K_n}{\log \delta_n} - \lim_{n\to\infty}\frac{1}{\log \delta_n}\log\sum_{k\in K_n}\mu^{-q}[B_{\delta_n}(k)] = \widetilde{\theta}(0) - \widetilde{\theta}(-q)$$

对 $q \in \mathcal{D}(C_0)$ 成立. □

4.3.2 推论——指数收敛

如果 $C_0(q)$ 满足广义假设 4.2.2, 则 $Y_n \xrightarrow{\exp} y_0 = C_0'(0)$. 进一步, $I_0(y_0) = 0$ 是 $I_0(y)$ 的唯一极小值.

证明 因为 $C_0(q)$ 满足广义假设 4.2.2, $0 \in \mathrm{int}\mathcal{D}(C_0)$. 同时, C_0 在 $\mathrm{int}\mathcal{D}(C_0)$ 是可微的, 所以 y_0 存在. 从而由定理 4.2.3 即得结论. □

4.3.3 例子

再次考虑第 3 章的多项分布测度的例子. 在此种情形中, 第 n 级的盒子或子区间可以很容易的用一个长度为 n 的数字串 ω 刻画, 即 $\omega \in \Omega^n$. $J_n(\omega)$ 是由 ω 刻画的子区间. 具有非零测度的子区间用 $\omega \in \Omega_0^n$ 刻画, 且具有宽度 $\delta_n = b^{-n}$. 所以, 对 $\omega = (\omega_1, \cdots, \omega_n) \in \Omega_0^n$, $\omega_i \in \Omega_0$ 有

$$Y_n(\omega) = \frac{\log \mu[J_n(\omega)]}{\log \delta_n} = \frac{\log\left(\prod_{i=1}^n p_{\omega_i}\right)}{\log b^{-n}} = -\frac{1}{n}\sum_{i=1}^n \log_b p_{\omega_i},$$

从而该例中的 Y_n 可以明确的写成一个部分和的平均, 它以指数级速率收敛于 y_0, 根据式 4.6,

$$y_0 = C_0'(0) = \frac{-1}{\sharp\Omega_0}\sum_{\omega\in\omega_0}\log_b p_\omega.$$

因而, 如果以相同权重抽样, 则 $Y_n \xrightarrow{\exp} y_0$, 于是满足弱大数定律. □

4.3.4 推论——大偏差的界

如果 $C_0(q)$ 满足广义假设 4.2.2, 则

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{\log \nu_0^{(n)}\{B_{\delta_n} : -\varepsilon < Y_n - y \leqslant \varepsilon\}}{-\log \delta_n} = -I_0(y), \quad y \in \mathrm{int}\mathcal{D}(I_0).$$

进一步, 如果 $-1 \in \mathcal{D}(C_0)$, 则

$$I_0(y) = -\widetilde{\theta}(0) - \widetilde{f}(y), \tag{4.7}$$

其中 $\widetilde{f}(y)$ 由式 (2.7) 给出.

证明 双重极限等式可由定理 4.2.4 得到. 式 (4.7) 可由式 (4.3) 及如下关系得到

$$\nu_0^{(n)}\{B_{\delta_n} : -\varepsilon < Y_n - y \leqslant \varepsilon\} = \frac{\sharp K_n(y,\varepsilon)}{\sharp K_n}, \tag{4.8}$$

其中 $K_n(y,\varepsilon)$ 由定义 2.5.4 给出. □

下面的推论给出了一些条件, 在这些条件下 μ 成为弱意义下的一个重分形测度 (定义 2.5.4).

4.3.5 推论——Legendre 变换对

给定 $C_0(q)$, 也即 $\widetilde{\theta}(0) - \widetilde{\theta}(-q)$, 满足广义假设 4.2.2, 且 $-1 \in \mathcal{D}(C_0)$, 则有

$$\widetilde{f}(y) = \inf_q \left\{qy - \widetilde{\theta}(q)\right\}, \tag{4.9}$$

以及

$$\widetilde{\theta}(q) = \inf_y \left\{qy - \widetilde{f}(y)\right\}. \tag{4.10}$$

证明 注意 $I_0(y) = \sup_q\{qy - C_0(q)\}$ 为 $C_0(q)$ 的 Legendre 变换. 因为 $C_0(q)$ 满足广义假设 4.2.2 以及 $-1 \in \mathcal{D}(C_0)$, 所以根据式 (4.6) 和 (4.7) 可得

$$-\widetilde{\theta}(0) - \widetilde{f}(y) = \sup_q\{qy - \widetilde{\theta}(0) + \widetilde{\theta}(-q)\},$$

于是式 (4.9) 成立. 类似地, 由式 (4.4), $C_0(q) = \sup_y\{qy - I_0(y)\}$, 利用同样的代换即得式 (4.10). □

启发式地讨论: 如果 $\delta_n^{-\widetilde{f}(y)}$ 大约是测度 μ 为 δ_n^y 阶的盒子的个数, 则

$$\sum_{k\in K_n} \mu^q[B_{\delta_n}(k)] \sim \sum_y \delta_n^{qy}\delta_n^{-\widetilde{f}(y)} \sim \delta_n^{\inf_y\{qy-\widetilde{f}(y)\}}.$$

于是

$$\widetilde{\theta}(q) = \inf_y \left\{qy - \widetilde{f}(y)\right\}.$$

当然这个关系没有这么简单, 它依赖于凸性的讨论. 在 $\widetilde{\theta}(q)$ 不可微的例子中, 本章中给出的 Legendre 变换关系是不成立的.

4.3.6 例子 (Holley 和 Waymire, 1992)

设 $\mathcal{X}=[0,1]\times[0,1]$. 首先, 用一个 $b\times b$ 的 δ_n-网格覆盖 $\mathcal{X}$. 在每步迭代中, 将每个盒子划分成 b^2 个更小的盒子. 在第 n 步中 $\delta_n=b^{-n}$, $\sharp K_n=b^{2n}$.

取测度 $\mu=\frac{1}{2}\lambda_1\times\delta_0+\frac{1}{2}\lambda_2$, 其中 λ_1 是 $[0,1]$ 上的一维 Lebesgue 测度, δ_0 是位于零处的 Dirac 原子测度, λ_2 是 $[0,1]\times[0,1]$ 上的 Lebesgue 测度. 则

$$\begin{aligned}\sum_{k\in K_n}\mu^q[B_{\delta_n}(k)]&=b^n\left(\frac{1}{2}b^{-n}+\frac{1}{2}b^{-2n}\right)^q+(b^{2n}-b^n)\left(\frac{1}{2}b^{-2n}\right)^q\\&=\left(\frac{1}{2}b^{-2n}\right)^q b^n(b^n+1)^q+b^n(b^n-1)\left(\frac{1}{2}b^{-2n}\right)^q.\end{aligned}$$

不难看出, 如果 $q<1$, 则这个求和在 $n\to\infty$ 时趋于无穷大. 下面来确定这个收敛速率.

有两种情况: $q\geqslant 1$ 和 $q\leqslant 1$. 首先考虑 $q\leqslant 1$, 有

$$\sum_{k\in K_n}\mu^q[B_{\delta_n}(k)]=\left(\frac{1}{2}b^{-2n}\right)^q b^{2n}\left[b^{n(q-1)}(1+b^{-n})^q+(1-b^{-n})\right],\tag{4.11}$$

所以方括号里的量在 $n\to\infty$ 时趋于 1. 从而,

$$\begin{aligned}\widetilde{\theta}(q)&=\lim_{n\to\infty}\frac{\log\sum\limits_{k\in K_n}\mu^q[B_{\delta_n}(k)]}{\log(b^{-n})}\\&=2(q-1)-\lim_{n\to\infty}\frac{1}{n}\log_b\left[b^{n(q-1)}(1+b^{-n})^q+(1-b^{-n})\right]\\&=2(q-1),\quad 当\ q\leqslant 1.\end{aligned}$$

对 $q\geqslant 1$, 重新整理式 (4.11) 得

$$\sum_{k\in K_n}\mu^q[B_{\delta_n}(k)]=\left(\frac{1}{2}b^{-2n}\right)^q b^{2n}b^{n(q-1)}\left[(1+b^{-n})^q+(1-b^{-n})b^{-n(q-1)}\right].$$

类似地

$$\begin{aligned}\widetilde{\theta}(q)&=\lim_{n\to\infty}\frac{\log\sum\limits_{k\in K_n}\mu^q[B_{\delta_n}(k)]}{\log(b^{-n})}\\&=(q-1)+\lim_{n\to\infty}\frac{1}{n}\log_b\left[(1+b^{-n})^q+(1-b^{-n})b^{-n(q-1)}\right]\\&=(q-1),\quad 当\ q\geqslant 1.\end{aligned}$$

于是有

$$\widetilde{\theta}(q)=\begin{cases}2(q-1), & 若\ q\leqslant 1,\\ q-1, & 若\ q\geqslant 1.\end{cases}$$

由式 (2.7) 可得

$$\widetilde{f}(y)=\begin{cases}1, & 若\ y=1,\\ 2, & 若\ y=2,\\ -\infty, & 其他,\end{cases}$$

以及

$$\dim_H\left\{x:\lim_{n\to\infty}\frac{\log\mu[B_{\delta_n}(x)]}{\log\delta_n}=y\right\}=\begin{cases}1, & 若\ y=1,\\ 2, & 若\ y=2,\\ 0, & 其他.\end{cases}$$

注意这里 $\widetilde{\theta}(q)$ 连续但不可微, 并且

$$\widetilde{f}(y)\neq\inf_q\left\{qy-\widetilde{\theta}(q)\right\}=\begin{cases}y, & 若\ 1\leqslant y\leqslant 2,\\ -\infty, & 其他.\end{cases}$$

□

4.4 样本测度组成的族

序列 Y_n 以指数速率收敛到的点依赖于样本测度. 通过考虑一族样本测度, 得到了一族极限点. 本节要刻画它们的关系, 这里没有要求用这些关系确定 μ 为弱意义下的重分形测度的条件 (见推论 4.3.5). 然而, 这些关系形成了一些想法的基础, 这些想法可以用来确定 μ 何时为强意义下的重分形测度.

4.4.1 定义

定义集函数 M_m

$$M_m(B)=\begin{cases}\mu^m(B), & 若\ \mu(B)\neq 0,\\ 0, & 若\ \mu(B)=0,\end{cases}\tag{4.12}$$

其中 $m\in\mathbb{R}$, $B\in\mathcal{B}(\mathcal{X})$. 于是样本测度定义为

$$\nu_m^{(n)}(A)=\frac{\displaystyle\sum_{j\in K_n}M_m[A\cap B_{\delta_n}(j)]}{\displaystyle\sum_{k\in K_n}\mu^m[B_{\delta_n}(k)]},\quad A\in\mathcal{F}_n,\tag{4.13}$$

其中 m 是一个数, 满足 $-m\in\mathrm{int}\mathcal{D}(C_0)$.

□

在这种情形, $C(q)$ 和 $I(y)$ 分别记为 $C_m(q)$ 和 $I_m(y)$. 在 $m=0$ 时, 这就是前一节考虑的均匀测度. 当 $m=1$ 时, 每个盒子分配的权重与构造过程中访问这个盒子的概率成正比. 在 $m<0$ 时, 盒子的样本测度为与构造过程中访问该盒子的累计次数成反比的概率.

4.4.2 定理——重整累积生成函数

如果 $-m \in \mathcal{D}(C_0)$, 则

$$C_m(q) = C_0(q-m) - C_0(-m), \tag{4.14}$$

因而, $\mathcal{D}(C_m) = \{q : C_0(q-m) < \infty\}$. 同时,

$$I_m(y) = C_0(-m) + my + I_0(y). \tag{4.15}$$

证明 在式 (4.2) 中代入 $\nu_m^{(n)}$ 并化简即得式 (4.14). 由式 (4.3) 和 (4.14) 可得

$$\begin{aligned} I_m(y) &= \sup_q \{qy - C_0(q-m) + C_0(-m)\} \\ &= C_0(-m) + \sup_q \{(q+m)y - C_0(q)\} \\ &= C_0(-m) + my + I_0(y). \end{aligned}$$

□

由式 (4.14) 和 (4.15) 可见: 通过移位型变换, $C_m(q)$ 和 $I_m(y)$ 分别和 $C_0(q)$ 和 $I_0(y)$ 联系起来. 进一步, 利用式 (4.6) 和 (4.7) 可以将 $C_m(q)$ 和 $I_m(y)$ 分别与 $\widetilde{\theta}(q)$ 和重分形谱 $\widetilde{f}(y)$ 联系起来.

4.4.3 引理

给定 $C_0(q)$ 满足广义假设 4.2.2 以及 $-m \in \mathrm{int}\mathcal{D}(C_0)$, 则 $C_m(q)$ 也满足广义假设 4.2.2. 进一步, $I_m(y)$ 在 $\mathcal{D}(I_0)$ 上连续.

证明 式 (4.14) 和 (4.15) 在 $-m \in \mathcal{D}(C_0)$ 时成立. 另一方面, 该假设要求 $0 \in \mathrm{int}\mathcal{D}(C)$, 而这个可由 $-m \in \mathrm{int}\mathcal{D}(C_0)$ 得到. $C_m(q)$ 的另一个需要满足的条件直接可由式 (4.14) 得到. $I_m(y)$ 的连续性可由式 (4.15) 和 I_0 的连续性得到. □

4.4.4 推论——指数收敛

如果 $C_0(q)$ 满足广义假设 4.2.2 以及 $-m \in \mathrm{int}\mathcal{D}(C_0)$, 则相对于 $\nu_m^{(n)}$, $Y_n \xrightarrow{\exp} y_m$, 其中 $y_m = C_m'(0) = C_0'(-m)$. 进一步, $I_m(y_m) = 0$ 是 $I_m(y)$ 的唯一极小值.

证明 可由定理 4.2.3 和引理 4.4.3 直接得到. □

4.4.5 大偏差的界

给定 $C_0(q)$ 满足广义假设 4.2.2 以及 $-m \in \mathrm{int}\mathcal{D}(C_0)$, 则

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{\log \nu_m^{(n)}\{B_{\delta_n}: -\varepsilon < Y_n - y \leqslant \varepsilon\}}{-\log \delta_n} = -I_m(y), \quad y \in \mathrm{int}\mathcal{D}(I_0).$$

进一步, 如果 $-1 \in \mathcal{D}(C_0)$, 则

$$I_m(y) = -\widetilde{\theta}(m) + my - \widetilde{f}(y). \tag{4.16}$$

证明　双重极限等式由定理 4.2.4 和引理 4.4.3 得到. 如果 $-1 \in \mathcal{D}(C_0)$, 则 $\widetilde{\theta}(0)$ 存在且有限, 于是由式 (4.15), (4.7) 和 (4.6) 即得式 (4.16).　□

4.4.6 注记

利用下面的近似估计可以粗略地得到式 (4.16):

$$\sum_{k\in K_n(y,\varepsilon)} \mu^m[B_{\delta_n}(k)] \approx \delta_n^{my}\sharp K_n(y,\varepsilon),$$

其中 $K_n(y,\varepsilon)$ 由定义 2.5.4 中给出. 于是,

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{\log \nu_m^{(n)}\{B_{\delta_n}: -\varepsilon < Y_n - y \leqslant \varepsilon\}}{-\log \delta_n}\\
&=\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{-\log\delta_n}\log\left(\frac{\displaystyle\sum_{k\in K_n(y,\varepsilon)}\mu^m[B_{\delta_n}(k)]}{\displaystyle\sum_{k\in K_n}\mu^m[B_{\delta_n}(k)]}\right)\\
&\approx\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{-\log\delta_n}\log(\delta_n^{my}\sharp K_n(y,\varepsilon)) + \widetilde{\theta}(m)\\
&= -my + \widetilde{f}(y) + \widetilde{\theta}(m).
\end{aligned}$$

□

4.4.7 推论

给定 $C_0(q)$ 满足广义假设 4.2.2 以及 $-m \in \mathrm{int}\mathcal{D}(C_0)$, 则 $y_m = C_m'(0) = C_0'(-m)$ 是唯一使得下确界 $\widetilde{\theta}(m) = \inf_y\{my - \widetilde{f}(y)\}$ 达到的 y 值. 进一步, y_m 看做 m 的函数对所有满足 $-m \in \mathrm{int}\mathcal{D}(C_0)$ 的 m 连续.

证明　y_m 的连续性可由 $C_0(q)$ 在 $\mathrm{int}\mathcal{D}(C_0)$ 上可微的假设条件得到. 根据推论 4.4.4 知

$$\inf_y I_m(y) = I_m(y_m) = 0,$$

其中 y_m 是唯一的. 由式 (4.16) 得

$$\inf_y I_m(y) = \inf_y \left\{ -\widetilde{\theta}(m) + my - \widetilde{f}(y) \right\} = 0,$$

从而结论得证. □

4.4.8 推论

给定 $C_0(q)$, 也即 $\widetilde{\theta}(0) - \widetilde{\theta}(-q)$, 满足广义假设 4.2.2, 且 $-1 \in \mathcal{D}(C_0)$, 则测度 μ 是满足弱意义下格公式的重分形测度 (见定义 2.5.4). 并且对于 $-q \in \mathrm{int}\mathcal{D}(C_0)$, $\widetilde{\theta}(q) = \inf_y\{qy - \widetilde{f}(y)\}$ 中的下确界在 $y_q = \widetilde{\theta}'(-q)$ 达到.

证明 第一个结论由推论 4.3.5 可得. 下确界的求解由式 (4.6) 和推论 4.4.7 可得. □

4.4.9 例子

考虑与例 4.3.3 中一样的情形, 只不过这里样本测度取 $\nu_m^{(n)}$. 这种情形具有一列极限点 y_m, 其中

$$y_m = C_m'(0) = \frac{-1}{\left(\sum_{i \in \Omega_0} p_i^q\right)} \sum_{\omega \in \Omega_0} p_\omega^q \log_b p_\omega,$$

这与第 3 章中的 y_q 一样, 那里用到了 Lagrange 乘子. □

到目前为止, 所得的结论似乎与弱大数定律 (见例 4.3.3) 类似. 已经证明在某些条件下, 譬如给定广义假设 4.2.2, $\widetilde{\theta}(q)$ 和 $\widetilde{f}(y)$ 可以通过 Legendre 变换联系起来, 也就是, 全局平均与测度 μ 的局部性态是相关的. 重分形谱 $\widetilde{f}(y)$ 是测度 μ 的局部性态的概率分布的幂律指数, 在这一意义下, 它是一维的. 然而, 它不一定是前面定义的集合的几何维数 (譬如: Hausdorff 维数、盒维数、填充维数). $\widetilde{f}(y)$ 作为一个 Hausdorff 维数的解释依赖于 $\nu_m^{(n)}$ 在 $\mathcal{B}(\mathcal{X})$ 上的扩充. 此时, 与 Hausdorff 维数相关的结论与强大数定律更为类似. 在下一节将进一步讨论这些概念.

4.5 Hausdorff 维数

我们已经用 $B_{\delta_n}(k)$ 表示格覆盖中的第 k 个盒子. 测度 $\nu_m^{(n)}$ 定义在由第 n 步中的盒子生成的 σ-子代数 $\mathcal{F}_n$ 上. 现在考虑点 $x \in \mathcal{X}$, 我们对包含这个点的盒子感兴趣. 这个点包含在指标为 $k(x)$ 的盒子中, 即 $B_{\delta_n}(k(x))$. 为了避免记号过于繁杂, 如果上下文不会引起混淆, 就简记为 $B_{\delta_n}(x)$. 进一步, 记 $\nu_m^{(\infty)}$ 为 $\nu_m^{(n)}$ 在 $n \to \infty$ 时的归纳极限测度.

4.5.1　定义

对 $y > 0$, 定义集合 $\widetilde{F}(y)$ 和 $\widetilde{F}_m(y)$ 如下

$$\widetilde{F}(y) = \left\{x \in \mathcal{X} : \lim_{n\to\infty} Y_n(x) = y\right\} = \left\{x \in \mathcal{X} : \lim_{n\to\infty} \frac{\log \mu[B_{\delta_n}(x)]}{\log \delta_n} = y\right\}, \quad (4.17)$$

以及

$$\widetilde{F}_m(y) = \left\{x \in \mathcal{X} : \lim_{n\to\infty} \frac{\log \nu_m^{(\infty)}[B_{\delta_n}(x)]}{\log \delta_n} = y\right\}. \qquad \square$$

本节估计 $\widetilde{F}(y)$ 的 Hausdorff 维数, 这包含几个步骤. 讨论基于 §4.4 中定义的样本测度 $\nu_m^{(n)}$ 的性态. 然而, 这里需要将该测度延拓到 $\mathcal{B}(\mathcal{X})$ 上. 作为一个推论, $Y_n \to y_m = C_0'(-m)$, $\nu_m^{(\infty)}$-a.s., 因而有与强大数定律更为类似的地方. 利用这一点, 可以证明集合 $\widetilde{F}(y_m)$ 和 $\widetilde{F}_m(my_m - \widetilde{\theta}(m))$ 在除去一个零测集后是相同的. 从而可以研究 $\nu_m^{(\infty)}$ 的局部性态. 接着利用 Billingsley 的一个定理确定 $\widetilde{F}_m(my_m - \widetilde{\theta}(m))$ 的 Hausdorff 维数. 这样便回到了我们关心的集合 $\widetilde{F}(y)$.

4.5.2　扩充定理

设 $\nu_m^{(n)}$ 是一致的, 即

$$\nu_m^{(n+1)}(E) = \nu_m^{(n)}(E), \quad \forall E \in \mathcal{F}_n, \quad n = 1, 2, \cdots,$$

以及

$$\nu_m^{(\infty)}(E) = \nu_m^{(n)}(E), \quad \forall E \in \mathcal{F}_n, \quad n = 1, 2, \cdots,$$

其中 $\nu_m^{(\infty)}$ 为 $\nu_m^{(n)}$ 在 $n \to \infty$ 时的归纳极限测度. 则 $\nu_m^{(\infty)}$ 是 $\nu_m^{(n)}$ 在 $\mathcal{B}([0,1]^d)$ 上的唯一扩充.

证明　易见 $E \in \mathcal{F}_n \Rightarrow E \in \mathcal{F}_{n+1}$. 根据 Kolmogorov 扩充定理 (Rényi, 1970, 286 页; Breiman, 1968, 24 页). $\square$

4.5.3　引理

设 $\nu_m^{(n)}$ 如定理 4.5.2 中所述是一致的, 则 $Y_n \to y_m$, $\nu_m^{(\infty)}$-a.s., 或者等价的有 $\nu_m^{(\infty)}[\widetilde{F}(y_m)] = 1$.

证明　由定理 4.5.2 知 $\nu_m^{(\infty)}$ 是 $\nu_m^{(n)}$ 在 $\mathcal{B}([0,1]^d)$ 上的唯一扩充, 再根据定理 B.3.15 可得 $Y_n \to y_m$, $\nu_m^{(\infty)}$-a.s. $\square$

4.5.4　引理

设 $\nu_m^{(n)}$ 如定理 4.5.2 中所述是一致的, 则

$$\widetilde{F}(y_m) \subseteq \widetilde{F}_m\left(my_m - \widetilde{\theta}(m)\right).$$

进一步, 不相交的部分为零测集, 即

$$\nu_m^{(\infty)}\left[\widetilde{F}_m(my_m-\widetilde{\theta}(m))\backslash\widetilde{F}(y_m)\right]=0.$$

证明 对所有 $x\in\widetilde{F}(y_m)$,

$$\begin{aligned}\lim_{n\to\infty}\frac{\log\nu_m^{(\infty)}[B_{\delta_n}(x)]}{\log\delta_n}&=\lim_{n\to\infty}\frac{\log\nu_m^{(n)}[B_{\delta_n}(x)]}{\log\delta_n}\\&=\lim_{n\to\infty}\left(\frac{\log\mu^m[B_{\delta_n}(x)]}{\log\delta_n}-\frac{\sum_k\log\mu^m[B_{\delta_n}(k)]}{\log\delta_n}\right)\\&=my_m-\widetilde{\theta}(m).\end{aligned}$$

所以 $\widetilde{F}(y_m)\subseteq\widetilde{F}_m\left(my_m-\widetilde{\theta}(m)\right)$. 第二个叙述由引理 4.5.3 可得. □

我们需要 Billingsley 的一个结论 (1965). 先给出 Hausdorff 维数的一个推广, 然后再给出 Billingsley 定理.

4.5.5 广义 Hausdorff 维数的定义 (Cutler, 1986)

设 $(\mathcal{X},\mathcal{B}(\mathcal{X}),\nu)$ 为一个概率空间, 其中 $\mathcal{X}\subseteq\mathbb{R}^d$, $\mathcal{B}(\mathcal{X})$ 是 $\mathcal{X}$ 的 Borel 子集构成的 σ-代数. 定义 $F\subset\mathcal{X}$ 的广义 s-维 Hausdorff 测度为

$$\mathcal{G}^s(F)=\lim_{\delta\to0}\inf_{\{U_i\}}\left\{\sum_{i=1}^{\infty}\nu(U_i)^s:\nu(U_i)\leqslant\delta\text{ 且 }F\subseteq\bigcup_{i=1}^{\infty}U_i\right\}.$$

相对于测度 ν 的广义 Hausdorff 维数 —— $\dim_\nu(F)$ 为

$$\dim_\nu(F)=\inf\{s:\mathcal{G}^s(F)=0\}=\sup\{s:\mathcal{G}^s(F)=\infty\}.$$ □

接下来的内容, 利用由 b-进制方体构成的格覆盖. 一个格覆盖中第 n 级的 d-维 b-进制方体的边界被长度为 n 的 b-进制展开式所决定, 不过这里每个边界的右边是开的.

4.5.6 Billingsley 定理 (Billingsley, 1965; Cutler, 1986, 1477 页)

设 λ 和 ν 是 $\mathcal{B}([0,1]^d)$ 上无原子的概率测度, 并且

$$G(y)\subseteq\left\{x:\lim_{n\to\infty}\frac{\log\nu[B_n(x)]}{\log\lambda[B_n(x)]}=y\right\},$$

其中 $B_n(x)$ 是包含点 x 体积为 b^{-nd} 的 d-维 b-进制方体. 则

$$\dim_\lambda G(y)=y\dim_\nu G(y).$$ □

4.5.7 推论 (Billingsley, 1965; Cutler, 1986, 1477 页)

设 λ 和 ν 是 $\mathcal{B}([0,1]^d)$ 上无原子的概率测度, $G(y)$ 如定理 4.5.6 中所述, 并且 $\nu[G(y)]>0$. 则 $\dim_\nu G(y)=1$. 进一步, 若 λ 表示 Lebesgue 测度, 则

$$\dim_H G(y)=d\dim_\lambda G(y)=dy. \qquad \square$$

4.5.8 引理

设 $\nu_m^{(n)}$ 如定理 4.5.2 中所述是一致的, $\nu_m^{(n)}$ 的归纳极限测度 $\nu_m^{(\infty)}$ 是 $\mathcal{B}([0,1]^d)$ 上无原子的概率测度. 则

$$\dim_H \widetilde{F}_m\left(my_m-\widetilde{\theta}(m)\right)=my_m-\widetilde{\theta}(m).$$

证明 在推论 4.5.7 中令 $\nu=\nu_m^{(\infty)}$. 据引理 4.5.4 有

$$\nu_m^{(\infty)}\left[\widetilde{F}_m(my_m-\widetilde{\theta}(m))\right]=1.$$

再注意到 $G(y/d)=\widetilde{F}_m(y)$ 即得结论. $\square$

4.5.9 定理

设 $\mathcal{X}=[0,1]^d$, $B_{\delta_n}(k)$ 为由宽度为 $\delta_n=b^{-n}$ 的 b-进制方体组成的系统, 其中 b 为一个 $\geqslant 2$ 的整数. 给定 $C_0(q)$ 满足广义假设 4.2.2, $-m\in \mathrm{int}\mathcal{D}(C_0)$ 以及 $-1\in\mathcal{D}(C_0)$, $\nu_m^{(n)}$ 如定理 4.5.2 中所述是一致的. 则

$$\dim_H \widetilde{F}(y_m)=\widetilde{f}(y_m),$$

其中 $y_m=C_m'(0)=C_0'(-m)$.

证明 据引理 4.5.4 和 4.5.8 有: $\dim_H \widetilde{F}(y_m)=my_m-\widetilde{\theta}(m)$. 由推论 4.4.4 得 $I_m(y_m)=0$, 从而由式 (4.16) 即得结论. $\square$

4.5.10 推论

设 $\mathcal{X}=[0,1]^d$, $B_{\delta_n}(k)$ 为由宽度为 $\delta_n=b^{-n}$ 的 b-进制方体组成的系统, 其中 $b\geqslant 2$ 为整数. 给定 $C_0(q)$ 满足广义假设 4.2.2, $-1\in\mathcal{D}(C_0)$, 并且对所有满足 $-m\in\mathcal{D}(C_0)$ 的 m, $\nu_m^{(n)}$ 如定理 4.5.2 中所述是一致的. 则

$$\dim_H \widetilde{F}(y)=\widetilde{f}(y), \quad 对 y\in(y_{\min},y_{\max}),$$

其中 $y_{\min}=\lim_{m\to q_{\max}} y_m$, $y_{\max}=\lim_{m\to q_{\min}} y_m$, $(q_{\min},q_{\max})=\mathcal{D}(C_0)$ 以及 $y_m=C_0'(-m)$.

证明 由定理 4.5.9 得: $\dim_H \widetilde{F}(y_m)=\widetilde{f}(y_m)$, 又因为 $y_m=C_0'(-m)$ 是关于 m 的连续函数, 并且可取到 $(y_{\min},y_{\max})$ 上的所有值, 所以结论得证. $\square$

4.5.11 例子——多项分布测度

把第 3 章中的多项分布测度放到这个理论框架中来, 即 $\mathcal{X}=[0,1]$ 以及 $\delta_n=b^{-n}$, 其中 $b\geqslant 2$ 为整数. 本章中用 $B_{\delta_n}(k)$(k 为一个任意的指标) 表示这些盒子. 在这个例子中, 这些盒子可以简单地用 $J_n(\omega)$ 表示, 其中 $\omega\in\Omega^n$, Ω^n 是由长度为 n 的序列组成的空间 (见第 3 章). 格系统 $J_n(\omega)$, $\omega\in\Omega^n$ 与 $B_{\delta_n}(k)$ 构成的格系统几乎是一样的, 除了 $J_n(\omega)$ 是闭区间, 而 $B_{\delta_n}(k)$ 是半开的. 通过指定 $B_{\delta_n}(k)$ 是半开的可以确保每个 $x\in\mathcal{X}$ 恰好属于唯一的某个盒子, 从而式 (4.17) 中 $\widetilde{F}(y)$ 的定义是合理的. 在这个多项分布测度的例子中, 一个数 x 的 b-进制展开式确定了它属于唯一的某个盒子. 在多项分布测度的情形中, 区间的边界具有零测度, 于是在两种覆盖体系下的函数 $\widetilde{\theta}(q)$ 和 $\widetilde{f}(y)$ 均是一样的.

在这个例子里, $\mathcal{F}_n$ 也可以定义为由没有重叠的区间 $J_n(\omega)$, $\omega\in\Omega^n$ 生成的子 σ-代数. 所以若 $A\in\mathcal{F}_n$, 则 $A\cap J_n(\omega)=\varnothing$ 或者 $A\cap J_n(\omega)=J_n(\omega)$. $\mathcal{F}_n$ 上的测度 μ 由式 (3.1) 定义. 根据构造过程, 易见它有一个到 $\mathcal{B}([0,1])$ 上的扩充.

现在考虑样本测度构成的族. 回忆 $\Omega_0=\{\omega\in\Omega:p_\omega>0\}$, 从而 $\Omega_0^n=\{\omega\in\Omega^n:\mu[J_n(\omega)]>0\}$. 设 $A\in\mathcal{F}_n$, 则 $A\in\mathcal{F}_{n+1}$, 于是

$$\begin{aligned}\nu_m^{(n+1)}(A)&=\frac{\displaystyle\sum_{\omega\in\Omega_0^{n+1}}M_m[A\cap J_{n+1}(\omega)]}{\left(\displaystyle\sum_{\omega\in\Omega_0}p_\omega^m\right)^{n+1}}\\&=\frac{\displaystyle\sum_{\omega\in\Omega_0^n}\sum_{\sigma\in\Omega_0}M_m[A\cap J_{n+1}(\omega\bowtie\sigma)]}{\left(\displaystyle\sum_{\omega\in\Omega_0}p_\omega^m\right)^{n+1}},\end{aligned}$$

其中 $\omega\bowtie\sigma$ 表示将 σ 串接在长度为 n 的序列 ω 后面, 集函数 M_m 由式 (4.12) 定义. 由于 $A\in\mathcal{F}_n$, 故对给定的 $\omega\in\Omega_0^n$, 如果对某个 $\sigma\in\Omega_0$ 有 $A\cap J_{n+1}(\omega\bowtie\sigma)\neq\varnothing$, 则对所有的 $\sigma\in\Omega_0$ 均有 $A\cap J_{n+1}(\omega\bowtie\sigma)\neq\varnothing$. 所以

$$\begin{aligned}\nu_m^{(n+1)}(A)&=\frac{\displaystyle\sum_{\omega\in\Omega_0^n}M_m[A\cap J_n(\omega)]\sum_{\sigma\in\Omega_0}p_\sigma^m}{\left(\displaystyle\sum_{\omega\in\Omega_0}p_\omega^m\right)^{n+1}}\\&=\nu_m^{(n)}(A).\end{aligned}$$

于是 $\nu_m^{(n)}$ 是一致的. 由定理 4.5.2 知, $\nu_m^{(n)}$ 有一个到 $\mathcal{B}([0,1])$ 上的扩充, 记为 $\nu_m^{(\infty)}$.

根据式 (3.8) 可得

$$\widetilde{\theta}(q) = -\log_b \sum_{\omega \in \Omega_0} p_\omega^q,$$

以及由式 (4.6) 得 $C_0(q) = \widetilde{\theta}(0) - \widetilde{\theta}(-q)$. 于是

$$I_0(y) = \sup_q \{qy - C_0(q)\} = \widetilde{D}_0 - \inf_q \left\{qy - \widetilde{\theta}(q)\right\}.$$

再由 $I_0(y)$ 的连续性有 $I_0(y) = \widetilde{D}_0 - \widetilde{f}(y)$, 进一步, $\dim_H \widetilde{F}(y) = \widetilde{f}(y)$. 式 (4.9) 和 (4.10) 中的 Legendre 变换关系如图 3.2~ 图 3.4 所示.

第 5 章　点中心情形的重分形

5.1 引　　言

本章按第 4 章的体系进行, 不同的是现在使用点中心构造. 即我们对函数 $f(y)$ 和 $\theta(q)$ 之间的关系感兴趣, 这里 $f(y)$ 和 $\theta(q)$ 分别由式 (2.9) 和 (2.5) 定义, 而且, 这时 $f(y)$ 可以看成 Hausdorff 维数.

在第 4 章, 用宽度为 δ_n(当 $n \to \infty$ 时, $\delta_n \to 0$) 的 d-维盒子覆盖集合 $\mathcal{X} \subset \mathbb{R}^d$ 的方式分析测度的行为. 第 n 步的盒子生成一子 σ 域 $\mathcal{F}_n$. $\mathcal{F}_n$ 上定义的样本测度间就会产生一些联系. 本章也会定义一族样本测度, 只不过这些测度都定义在同一测度空间 $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ 上.

§5.2 对一般样本测度定义大偏差体系; §5.3 给出一族样本测度的定义; §5.4 给出 $f(y)$ 可以解释为 Hausdorff 维数的条件. 这 3 节和第 4 章的讨论是平行的. 假设读者对第 4 章的内容是熟悉的, 因而涉及第 4 章的相关内容时不再重复.

§5.5 对格和点中心构造之间的联系进行了研究. 其中以下问题是让人感兴趣的: 什么时候 Rényi 维数和重分形谱在两种构造中是一致的? 两种构造的另一个不同之处是极限的存在性、覆盖格或球的宽度. 这些问题都会在 §5.5 进行讨论.

5.2 大偏差体系

考虑概率空间 $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \upsilon^{(\delta)})$, 这里 $\mathcal{B}(\mathcal{X})$ 是 $\mathcal{X} \subset \mathbb{R}^d$ 的 Borel 子集, $\nu^{(\delta)}$ 是某个一般样本测度, 它是 δ 的函数. 注意这和 §4.2 基于格的构造是不同的, 那里涉及一列概率空间 $(\mathcal{X}, \mathcal{F}_n, \nu^n)$, 这里只用到一个概率空间 $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \nu^\delta)$, 但是概率测度 $\nu^{(\delta)}$ 是随着 δ 变动的.

定义测度函数族 $\{U_\delta\}_\delta$ 如下:

$$U_\delta : (\mathcal{X}, \mathcal{B}(\mathcal{X})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})),$$

$$U_\delta(x) = \begin{cases} -\log \mu[S_\delta(x)], & \mu[S_\delta(x)] > 0, \\ 0, & \mu[S_\delta(x)] = 0, \end{cases}$$

这里 $S_\delta(x)$ 是中心为 x 半径为 δ 的 d 维闭球. 同样地, 定义函数 $\{Y_\delta\}_{\delta>0}$ 为

$$Y_\delta(x) = \frac{U_\delta(x)}{-\log \delta}.$$

我们对 Y_δ 在概率空间 $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \nu^{(\delta)})$ 上当 $\delta \to 0$ 时的行为感兴趣.

令 $\mathcal{X}_\delta = \{x \in \mathcal{X} : \mu[S_\delta(x)] > 0\}$. $\mathcal{X}_\delta$ 有时被称为 μ 的支撑的 δ-平行体或香肠集. 由 Falconer (1990, 命题 3.2) 有

$$\widetilde{D}_0 = \dim_B \operatorname{supp}(\mu) = d - \lim_{\delta \to 0} \frac{\log \operatorname{Vol}^d(\mathcal{X}_\delta)}{\log \delta}.$$

以上假设极限存在, 这里 $\operatorname{Vol}^d(\mathcal{X}_\delta)$ 表示 $\mathcal{X}_\delta$ 的 d-维体积.

所有形如 $C(q), I(y), C_m(q)$ 和 $I_m(y)$ 的记号都是指本章定义的, 不同于第 4 章的定义.

5.2.1 重整累积生成函数

由 B.5 知, 重整累积生成函数为 $C(q), q \in \mathbb{R}$, 假如以下极限存在:

$$C(q) = \lim_{\delta \to 0} \frac{\log E[e^{qU_\delta(x)}]}{-\log \delta},$$

这里期望是对样本测度 ν^δ 取的. 因此,

$$C(q) = \lim_{\delta \to 0} \frac{\log \int_{\mathcal{X}_\delta} \mu^{-q}[S_\delta(x)] \nu^\delta(dx)}{-\log \delta}, \tag{5.1}$$

这里允许 $+\infty$ 为极限. 进一步, 令

$$I(y) = \sup_{q \in \mathbb{R}} \{qy - C(q)\}, \quad y > 0. \tag{5.2}$$

$I(y)$ 可以取 $+\infty$, 称为 $C(q)$ 的 Legendre 变换.

5.2.2 广义假设

1) $C(q)$ 满足假设 B.3.8.

2) $C(q)$ 在集合 $\mathcal{D}(C) = \{q \in \mathbb{R} : C(q) < \infty\}$ 的内部可微, 并且是急剧升降的 (见定义 B.2.1).

5.2.3 定理——指数收敛

假如 $C(q)$ 满足广义假设 5.2.2, 那么 $I(y)$ 在 $y_0 = C'(0)$ 处有唯一的极小值使得 $I(y_0) = 0$. 进一步, $Y_\delta \xrightarrow{\exp} y_0$ 相对于 ν^δ(见定义 B.3.13). 并且

$$C(q) = \sup_y \{qy - I(y)\}. \tag{5.3}$$

证明 在广义假设 5.2.2 下, 用定理 4.2.3 的证明方法可以很快得到该定理的证明. □

5.2.4　定理——大偏差的界

假如 $C(q)$ 满足广义假设 5.2.2, 那么

$$\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\frac{\log \nu^\delta\{x: -\varepsilon < Y_\delta(x) - y \leqslant \varepsilon\}}{-\log\delta} = -I(y),\ 其中\ y\in \mathrm{int}\mathcal{D}(I). \tag{5.4}$$

证明　在广义假设 5.2.2 下, 用定理 4.2.2 的证明方法可以很快得到上面定理的证明. □

5.3　一族样本测度

同 §4.4 中的定义方式一样, 可以定义一族样本测度. 从而产生依赖于样本测度的不同极限.

5.3.1　定义

定义一族样本测度 ν_m^δ 如下:

$$\nu_m^\delta(E) = \frac{\displaystyle\int_{E\cap\mathcal{X}_\delta}\mu^{m-1}[S_\delta(x)]\mu(dx)}{\displaystyle\int_{\mathcal{X}_\delta}\mu^{m-1}[S_\delta(x)]\mu(dx)}, \tag{5.5}$$

这里 $m\in\mathbb{R}$, δ 固定, $E\in\mathcal{B}(\mathcal{X})$.

在这种情况下, $C(q)$ 和 $I(y)$ 分别记作 $C_m(q)$ 和 $I_m(q)$. 注意到 ν_m^δ 是定义在 $\mathcal{B}(\mathcal{X})$ 上的, 而在 §4.4 中是定义在子 σ-域上的.

5.3.2　定理——重整累积生成函数

假如 $\theta(0)$ 存在, 或者等价地, $-1\in\mathcal{D}(C_0)$, 那么

$$C_0(q) = \theta(0) - \theta(-q),$$

这里 $\theta(q)$ 由定义 2.4.1 给出. 进一步, 如果 $-m\in\mathcal{D}(C_0)$, 那么

$$C_m(q) = C_0(q-m) - C_0(-m), \tag{5.6}$$

并且

$$I_m(y) = C_0(-m) + my + I_0(y). \tag{5.7}$$

证明　将 ν_m^δ 带入 (5.1) 式, 然后令 $m=0$, $q=-1$, 可得到 $C_0(-1)=\theta(0)$, 即如果 $-1\in\mathcal{D}(C_0)$, 则 $\theta(0)$ 存在. 因此, 假如 $-1\in\mathcal{D}(C_0)$, 那么 $C_0(q)=\theta(0)-\theta(-q)$. 同样地, 如果 $-m\in\mathcal{D}(C_0)$, 将 ν_m^δ 带入 (5.1) 式整理可得到 (5.6) 式. 利用证明定理 4.4.2 的方法可以证明 (5.7) 式. □

5.3.3 引理

假如 $C_0(q)$ 满足广义假设 5.2.2, 并且 $-m \in \mathrm{int}\mathcal{D}(C_0)$, 那么 $C_m(q)$ 满足广义假设 5.2.2. 进一步, $I_m(y)$ 是 $\mathcal{D}(I_0)$ 上的连续函数.

证明 证明方法同引理 4.4.3. □

5.3.4 推论——指数收敛

假如 $C_0(q)$ 满足广义假设 5.2.2, 并且 $-m \in \mathrm{int}\mathcal{D}(C_0)$, 那么, $Y_\delta \xrightarrow{\exp} y_m$ 相对于 ν_m^δ, 这里 $y_m = C_m'(0) = C_0'(-m)$. 进一步, $I_m(y_m)$ 是 $I_m(y)$ 的唯一极小值.

证明 由定理 5.2.3 和引理 5.3.2 即得. □

5.3.5 注记

假如 $\theta(0)$ 存在 (式 (2.5)), 那么由式 (2.9) 可得

$$f(y) = -\theta(0) + \lim_{\varepsilon\to 0}\lim_{\delta\to 0} \frac{\log \nu_0^{(\delta)}\{x \in \mathcal{X} : -\varepsilon < Y_\delta(x) - y \leqslant \varepsilon\}}{-\log\delta}, \tag{5.8}$$

当 $\nu_0^{(\delta)}\{x \in \mathcal{X} : -\varepsilon < Y_\delta(x) - y \leqslant \varepsilon\} = 0$ 时, 允许 $f(y) = -\infty$.

5.3.6 推论——大偏差界

假如 $C_0(q)$ 满足广义假设 5.2.2, 那么对满足 $-m \in \mathrm{int}\mathcal{D}(C_0)$ 的 m 和所有的 $y \in \mathrm{int}\mathcal{D}(I_0)$, 有

$$\lim_{\varepsilon\to 0}\lim_{\delta\to 0} \frac{\log \nu_m^{(\delta)}\{x \in \mathcal{X} : -\varepsilon < Y_\delta(x) - y \leqslant \varepsilon\}}{-\log\delta} = -I_m(y).$$

进一步, 假如 $-1 \in \mathcal{D}(C_0)$, 那么

$$I_m(y) = -\theta(m) + my - f(y), \tag{5.9}$$

这里 $\theta(m)$ 和 $f(y)$ 分别由式 (2.5) 和 (2.9) 定义.

证明 定理 5.2.4 保证了二重极限的存在. 假如 $-1 \in \mathcal{D}(C_0)$, 那么 $\theta(0)$ 存在并且有限. 联合式 (5.8) 可得 $I_0(y) = -f(y) - \theta(0)$. 然后带入式 (5.7) 得到式 (5.9). □

5.3.7 推论

假如 $C_0(q)$ 满足广义假设 5.2.2, $-m \in \mathrm{int}\mathcal{D}(C_0)$, 那么 $y_m = C_m'(0) = C_0'(-m)$ 是唯一使得 $\theta(m) = \inf_y\{my - f(y)\}$ 中下确界可达的点. 进一步, y_m 是关于 m 的连续函数.

证明 证明方法同推论 4.4.7. □

5.3.8 推论

假如 $C_0(q)$ 满足广义假设 5.2.2, $-1 \in \mathcal{D}(C_0)$, 那么测度 μ 是在较弱意义 (定义 2.5.5) 下满足点中心体系的重分形测度. 进一步, 式 $\theta(q) = \inf_y\{qy - f(y)\}$ 的下确界在值 $y = \theta'(-q), -q \in \mathrm{int}\mathcal{D}(C_0)$ 处取到.

证明 证明方法同推论 4.4.8. □

5.4 Hausdorff 维数

在 §4.5 讨论了在什么条件下, $\widetilde{f}(y)$ 可以解释成 $\widetilde{F}(y)$ 的 Hausdorff 维数. 本节将按同样的思路进行. 在 §4.5 借助了 Billingsley (1965) 的一个结果. 本节与之对应的是 Yong (1982) 的一个结果.

5.4.1 定义

对 $y > 0$, 定义集合 $F(y)$ 和 $F_m(y)$ 如下:

$$F(y) = \left\{x \in \mathcal{X} : \lim_{\delta\to 0} Y_\delta = y\right\} = \left\{x \in \mathcal{X} : \lim_{\delta\to 0} \frac{\log \mu[S_\delta(x)]}{\log \delta} = y\right\} \tag{5.10}$$

和

$$F_m(y) = \left\{x \in \mathcal{X} : \lim_{\delta\to 0} \frac{\log \nu_m^{(0^+)}[S_\delta(x)]}{\log \delta} = y\right\},$$

这里 $\nu_m^{(0^+)}$ 是 $\nu_m^{(\delta)}$ 当 $\delta \to 0$ 时的极限测度.

我们对 Y_δ 在概率空间 $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \nu_m^{(0^+)})$ 上的行为感兴趣. 在较好的条件下, 可以证明 $F(y_m) \subset F_m(my_m - \theta(m))$. 为此, 一些正则性条件是必须的.

5.4.2 引理

假设 $C_0(q)$ 满足广义假设 5.2.2, $-1 \in \mathrm{int}\mathcal{D}(C_0)$. 进一步地, 对任意满足 $-m \in \mathrm{int}\mathcal{D}(C_0)$ 的 m, 有 $\nu_m^{(0^+)}[F(y_m)] = 1$, 并且以下两条对任意的 $x \in F(y_m)$ 成立:

$$\lim_{\delta\to 0} \frac{\log \int_{S_\delta(x)} \mu^{m-1}[S_\delta(x)]\mu(dz)}{\log \delta} = \lim_{\delta\to 0} \frac{\log \mu^m[S_\delta(x)]}{\log \delta} \tag{5.11}$$

和

$$\lim_{\delta\to 0} \frac{\log \nu_m^{(0^+)}[S_\delta(x)]}{\log \delta} = \lim_{\delta\to 0} \frac{\log \nu_m^{(\delta)}[S_\delta(x)]}{\log \delta}. \tag{5.12}$$

那么, $F(y_m) \subset F_m(my_m - \theta(m))$, 这里 $y_m = C_m'(0) = C_0'(-m)$. 进一步, 不交的部分是零测集, 即

$$\nu_m^{(0^+)}[F_m(my_m - \theta(m)) \setminus F(y_m)] = 0.$$

证明　因为 $-1 \in \mathrm{int}\mathcal{D}(C_0)$, 所以 $\theta(0)$ 存在. 同样, 因为 $-m \in \mathrm{int}\mathcal{D}(C_0)$, 所以 $\theta(m)$ 和 y_m 存在. 考虑 $x \in F(y_m)$ 的点,

$$\lim_{\delta\to 0}\frac{\log \nu_m^{(0^+)}[S_\delta(x)]}{\log\delta} = \lim_{\delta\to 0}\frac{\log\displaystyle\int_{S_\delta(x)}\mu^{m-1}[S_\delta(x)]\mu(dz)}{\log\delta} - \theta(m)$$
$$= my_m - \theta(m).$$

因此, $F(y_m) \subset F_m(my_m - \theta(m))$. 注意到 $\nu_m^{(0^+)}[F(y_m)] = 1$, 第二个论断是显然的. □

注意到, 以上结果和引理 4.5.4 是不同的. 后者是基于格的构造, 给定相容和扩张定理 4.5.2, $\nu_m^{(\infty)}[F(y_m)] = 1$ 可由定理 B.3.15 得到. 在点中心构造情形, 式 (5.12) 是比相容和扩张定理 4.5.2 弱的条件, 因而仍要求 $\nu_m^{(0^+)}[F(y_m)] = 1$. 式 (5.11) 是某种光滑条件.

5.4.3　定理 (Young, 1982)

设 ν 是 $\mathbb{R}^d$ 上一有限非原子 Borel 测度, $F \subset \mathbb{R}^d$ 是 ν-可测的, 并且 $\nu(F) > 0$. 假设对任意的 $x \in F$,

$$\alpha_1 \leqslant \liminf_{\delta\to 0}\frac{\log\nu[S_\delta(x)]}{\log\delta} \leqslant \limsup_{\delta\to 0}\frac{\log\nu[S_\delta(x)]}{\log\delta} \leqslant \alpha_2.$$

那么, $\alpha_1 \leqslant \dim_H \leqslant \alpha_2$.

5.4.4　推论

设 $\nu_m^{(0^+)}$ 是有限非原子的 Borel 测度, $\nu_m^{(0^+)}[F_m(my_m - \theta(m))] > 0$. 那么对满足 $-m \in \mathrm{int}\mathcal{D}(C_0)$ 的 m, 有

$$\dim_H F_m(my_m - \theta(m)) = my_m - \theta(m).$$

证明　由定理 5.4.3 即得. □

5.4.5　定理

假设引理 5.4.2 的条件满足, 并且$\nu_m^{(0^+)}$是非原子的概率测度. 那么 $\dim_H F(y_m) = f(y_m)$, 这里 $y_m = C'_m(0) = C'_0(-m)$.

证明　由引理 5.4.2, $F(y_m) \subset F_m(my_m - \theta(m))$. 由假设可得 $\nu_m^{(0^+)}[F_m(my_m - \theta(m))] = 1$, 因此由推论 5.4.4 可得

$$\dim_H F(y_m) = my_m - \theta(m).$$

由于 $I_m(y_m) = 0$, 代入式 (5.9) 可得 $f(y_m) = my_m - \theta(m)$. □

5.4.6 推论

假设引理 5.4.2 的条件都满足, 并且 $\nu_m^{0^+}$ 是非原子的概率测度. 那么

$$\dim_H F(y) = f(y), \quad 对\ y \in (y_{\min}, y_{\max}),$$

这里, $y_{\min} = \lim\limits_{m \to q_{\max}} y_m, y_{\max} = \lim\limits_{m \to q_{\min}} y_m, (q_{\min}, q_{\max}), y_m = C_0'(-m)$.

证明 注意到 $y_m = C_0'(-m)$ 是 m 取值于 $(y_{\min}, y_{\max})$ 的一个连续函数, 由定理 5.4.5 可得到以上结果. □

5.5 格构造和点中心构造之间的关系

利用针对格和中心构造的样本函数族, 得到了熵函数和重整生成函数之间相同的关系. 保证 μ 在较弱意义下 (即 Legendre 变换关系成立) 是一个重分形测度的结果可由 Ellis (1984, 1985) 的大偏差结果得到, 这些结果在某种意义下和弱大数定理类似. 在格构造和中心构造的情形, 函数 $\widetilde{f}(y)$ 和 $f(y)$ 可以看做一种维数, 它们分别描叙了概率测度 ν_m^n 和 ν_m^δ 收敛的幂定律. 然而, 牵涉到 Hausdorff 维数的一些更强的结果就要求 Y_n 和 Y_δ 几乎必然收敛. 因此, 它们就更像强大数定理了.

本节想阐述以下 4 个问题:

1. 在什么条件下, 两种构造中划分集合的局部性质是相同的; 即什么时候, $F(y) = \widetilde{F}(y)$? 这里 $F(y)$ 和 $\widetilde{F}(y)$ 分别由式 (5.10) 和 (4.17) 给出.

2. 在什么条件下, Rényi 维数在两种构造中是相同的; 即什么时候 $D_q = \widetilde{D}_q$ 或 $\theta(q) = \widetilde{\theta}(q)$? 这里 $\theta(q)$ 和 $\widetilde{\theta}(q)$ 分别由式 (2.5) 和 (2.5) 给出.

3. 在格构造和中心构造中, 关于 δ 的极限的本质差别是什么? 即 $\delta_n \to 0$ 和 $\delta \to 0$ 的比较.

4. 在什么条件下重分形谱是一样的? 即什么时候 $f(y) = \widetilde{f}(y)$? 这里 $f(y)$ 和 $\widetilde{f}(y)$ 分别由式 (2.9) 和 (2.7) 给出.

关于以上问题, 虽然有些例子表明这些关系是成立的, 但是没有一个问题是完全可以解决的. 在一般情形可能需要更多的条件.

什么时候 $F(y) = \widetilde{F}(y)$?

以下结果是 Cawley 和 Mauldin (1992) 的结果应用到例 4.5.11 中多项测度 (multinomial measures) 的一个特殊情况. 这一结果对一般的莫朗级联 (将在第 6 章讨论) 情形也成立. 以下记号 (即, $\Omega, \Omega_0, \Omega^n, \Omega^\infty$ 等) 和例 4.5.11 相同. 另外, 假若 $\omega = (\omega_1, \omega_2, \cdots) \in \Omega^\infty$, 那么 $\omega|n$ 表示 $(\omega_1, \omega_2, \omega_n)$. 符号 $\bowtie$ 表示连接, 例如 $(\omega_1, \omega_2, \omega_n) \bowtie t = (\omega_1, \omega_2, \omega_n, t)$.

5.5.1　定理 (Cawley 和 Mauldin, 1992)

令 $\omega=(\omega_1,\omega_2,\cdots)\in\Omega^\infty$,

$$x=\sum_{k=1}^{\infty}\frac{\omega_k}{b^k},$$

$S_\delta(x)$ 表示中心在 x、半径为 δ 的闭球. 并且, μ 是基为 b 的多项测度 (multinomial measures)(定义见例 4.5.11). 那么

$$\limsup_{\delta\to 0}\frac{\log\mu[S_\delta(x)]}{\log\delta}\leqslant\limsup_{n\to\infty}\frac{\log\mu[J_n(\omega|n)]}{\log\delta_n},$$

这里 $\delta_n=b^{-n}$, 并且

$$J_n(\omega|n)=\Big[\sum_{j=1}^{n}\frac{\omega_j}{b^k},\frac{1}{b^n}+\sum_{j=1}^{n}\frac{\omega_j}{b^k}\Big].$$

证明　暂时固定 $\delta<1/(2b)$. 选取 n 使得 $J_n(\omega|n)\subset S_\delta(x)$, 但 $J_{n-1}(\omega|n-1)\not\subset S_\delta(x)$. 因此, $|J_n(\omega|n)|\leqslant 2\delta$. 由构造, $J_{n-1}(\omega|(n-1))\not\subset S_\delta(x)$, 因此, $J_{n-1}(\omega|(n-1))\geqslant\delta$. 故

$$\delta\leqslant|J_{n-1}(\omega|(n-1))|=b|J_n(\omega|(n))|\leqslant 2b\delta<1.$$

从而

$$\begin{aligned}\frac{\log\mu[S_\delta(x)]}{\log\delta}&\leqslant\frac{\log\mu[J_n(\omega|n)]}{\log\delta}\\&\leqslant\frac{\log\mu[J_n(\omega|n)]}{\log|J_n(\omega|n)|+\log b}\\&=\frac{\log\mu[J_n(\omega|n)]}{\log|J_n(\omega|n)|\left(1+\dfrac{\log b}{\log|J_n(\omega|n)|}\right)}\\&\leqslant\frac{\log\mu[J_n(\omega|n)]}{\log|J_n(\omega|n)|}\left(1-\frac{\log b}{\log|J_n(\omega|n)|}\right)\\&\leqslant\frac{\log\mu[J_n(\omega|n)]}{\log|J_n(\omega|n)|}\left(1-\frac{\log b}{\log 2\delta}\right).\end{aligned}$$

两边取极限即可得.　□

5.5.2　定理 (Cawley 和 Mauldin, 1992)

令 $\omega=(\omega_1,\omega_2,\cdots)\in\Omega^\infty$,

$$x=\sum_{k=1}^{\infty}\frac{\omega_k}{b^k},$$

$S_\delta(x)$ 表示中心在 x、半径为 δ 的闭球. 设 μ 为以 b 为基, 以 $p_0, \cdots, p_{b-1}$ 为权的多项分布测度 (见例 4.5.11). 假设对任意 $p_i > 0$, $p_j > 0 (i < j)$, 存在 $i < k < j$ 使得 $p_k = 0$. 那么

$$\liminf_{\delta\to 0} \frac{\log \mu[S_\delta(x)]}{\log \delta} \geqslant \liminf_{n\to\infty} \frac{\log \mu[J_n(\omega|n)]}{\log \delta_n},$$

这里 $\delta_n = b^{-n}$.

证明 令 $\mathcal{K} = \bigcap_{n=1}^{\infty} \bigcup_{\omega' \in \Omega_0^n} J_n(\omega')$, 这里 $\bigcup_{\omega' \in \Omega_0^n} J_n(\omega')$ 是具有正 μ 测度的 n-阶基本区间. 暂时固定 $\delta < 1/(2b)$. 令

$$h_\delta(x) = \max\{n : S_\delta(x) \cap \mathcal{K} \subset J_n(\omega|n)\},$$

这里 $\omega = (\omega_1, \omega_2, \cdots)$. 因此, $\log \mu[S_\delta(x)] \leqslant \log \mu[J_{h_\delta(x)}(\omega|h_\delta(x))]$. 选 $t \neq \omega_{h_\delta(x)+1}$ (即不等于 ω 中第 $h_\delta(x)+1$ 个数字) 和 $y \in J_{h_\delta(x)+1}(\omega|h_\delta(x)) \bowtie t \cap \mathcal{K} \cap S_\delta(x)$, 使得 $x \in J_{h_\delta(x)+1}(\omega|h_\delta(x)+|) \subset J_{h_\delta(x)}(\omega|h_\delta(x))$, 且

$$J_{h_\delta(x)+1}(\omega|h_\delta(x)+|) \cap J_{h_\delta(x)+1}(\omega|h_\delta(x)) \bowtie t = \varnothing.$$

因此,

$$\frac{1}{b} > 2\delta > |x-y|.$$

然而, 包含 x 和 y 的 b-进制区间之间有宽为 $b^{-(h_\delta(x))+1}$ 的间隙. 从而, $2\delta > |x-y| > b^{-(h_\delta(x))+1} = b^{-1}|J_{h_\delta(x)}(\omega|)|h_\delta(x))|$. 因此, $|J_{h_\delta(x)}(\omega|)|h_\delta(x))| < 2\delta b < 1$, 故

$$\begin{aligned}
\frac{\log \mu[S_\delta(x)]}{\log \delta} &\geqslant \frac{\log \mu[J_{h_\delta(x)}(\omega|h_\delta(x))]}{\log \delta} \\
&\geqslant \frac{\log \mu[J_{h_\delta(x)}(\omega|h_\delta(x))]}{\log |J_{h_\delta(x)}(\omega|h_\delta(x))| - \log(2b)} \\
&= \frac{\log \mu[J_{h_\delta(x)}(\omega|h_\delta(x))]}{\log |J_{h_\delta(x)}(\omega|h_\delta(x))| \left(1 - \dfrac{\log(2b)}{\log |J_{h_\delta(x)}(\omega|h_\delta(x))|}\right)} \\
&\geqslant \frac{\log \mu[J_{h_\delta(x)}(\omega||h_\delta(x))]}{\log |J_{h_\delta(x)}(\omega|h_\delta(x))|} \left(1 + \frac{\log(2b)}{\log |J_{h_\delta(x)}(\omega|h_\delta(x))|}\right).
\end{aligned}$$

然后取极限就可以得到所需结果. □

5.5.3 推论

设 μ 是满足定理 5.5.2 条件的多项分布测度 (例 4.5.11), $B_{\delta_n}(x)$ 是包含 x 的长度为 $\delta_n = b^{-n}$ 的 b-进制区间. 那么, $x \in F(y)$ 当且仅当 $x \in \widetilde{F}(y)$, 这里 $F(y)$ 和 $\widetilde{F}(y)$ 分别由方程 (4.17) 和 (5.10) 定义. 因此, $\dim_H F(y) = \dim_H \widetilde{F}(y) = \widetilde{f}(y)$.

证明　由定理 5.5.1 和 5.5.2 知, $F(y) = \widetilde{F}(y)$. 由例 4.5.11 知, $\dim_H F(y) = \dim_H \widetilde{F}(y) = \widetilde{f}(y)$. □

注意到以上例子要求有“间隙”(见定理 5.5.2). 同时注意到在点中心情形, 极限对所有的序列 $(\delta \to 0)$, 然而, 在格情形下, 需要特殊的序列 $\{\delta_n\}$.

什么时候 $D_q = \widetilde{D}_q$?

在例 3.5.1 中, 对于 Cantor 测度, $D_2 = \widetilde{D}_2$, 或者等价地 $\theta(2) = \widetilde{\theta}(2)$. 那里使用的方法对 $q = 3, 4, \cdots$ 同样适用.

5.5.4　定理

设 μ 是定理 5.5.2 中带有间隙的多项分布测度. 那么, 对 $q = 2, 3, \cdots$, $D_q = \widetilde{D}_q$, 或者等价地 $\theta(q) = \widetilde{\theta}(q)$. 进一步, $\theta(q)$ 中的极限 $(\delta \to 0)$ 对所有的序列都成立.

证明　证明类似于例 3.5.1. 设 $X_1, X_2, \cdots, X_q$ 是从 $\mathcal{K}_n$ 上的分布 (μ_n) 中选取的独立随机变量 (见例 3.5.1 中的定义). 定义

$$Y_n = \max\{|X_1 - X_q|, |X_2 - X_q|, \cdots, |X_{q-1} - X_q|\},$$

那么, 由 §2.4.5,

$$\int_{\mathcal{X}_n} \mu_n[S_\delta(x)]\mu_n(dx) = \Pr\{Y_n \leqslant \delta\}.$$

当 $n = 0$ 时, μ_0 就是单位区间上的一致分布. 正如例 3.5.1, 当 $X_1, X_2, \cdots, X_q$ 是一致分布的样本时, $g_0(y)$ 是 Y_0 的密度. 证明过程中 $g_0(y)$ 是什么形式不重要. 第一次划分的概率分布为 μ_1, 即第一阶的子区间赋予概率 p_1, 第二阶概率为 p_2 等等. 然而, 至少有一个替换区间具有零测度, 即具有间隔至少为子区间宽度的间隙. 因此, 假如 $Y_1 < b^{-1}$, 那么所有的 q 点 $X_1, X_2, \cdots, X_q$ 一定是同一区间的样本. 由于每一个子区间具有总质量等于 $p_0, \cdots, p_{q-1}$ 之一的一致分布, 所以

$$g_1(y) = b(p_0^q + \cdots + p_{b-1}^q)g_0(by),$$

这里 $0 \leqslant y \leqslant b^{-1}$. 在每次划分中都进行同样的讨论, 一般地,

$$g_n(y) = b(p_0^q + \cdots + p_{b-1}^q)g_{n-1}(by),$$

这里 $0 \leqslant y \leqslant b^{-n}$. 因此, 概率分布函数满足以下递推关系:

$$G_n(y) = (p_0^q + \cdots + p_{b-1}^q)G_{n-1}(by) = (p_0^q + \cdots + p_{b-1}^q)^n G_0(b^n y),$$

这里 $0 \leqslant y \leqslant b^{-n}$. 剩下的讨论和例 3.5.1 一样. □

注意到定理 5.5.4 和推论 5.5.3 都要求测度支撑在有间隙的集合上. 这个条件是充分的, 但是目前不知道它是否是必要的.

Mikosch 和 Wang (1993, 命题 2.1) 在一般情形下给出 $q>1$ 时 $\theta(q)=\widetilde{\theta}(q)$ 的条件. 他们要求测度 μ 具有紧支撑和连续性.

Riedi (1995, 命题 20) 证明了假如 μ 是任意的 Borel 测度, 那么, 对 $q>1$ 有

$$D_q=\frac{\widetilde{\theta}_R(q)}{q-1},$$

这里 $\widetilde{\theta}_R(q)$ 的定义如 §2.6.3. 回忆 $\widetilde{\theta}_R(q)$ 和 $\widetilde{\theta}(q)$ 有两点不同: $\widetilde{\theta}_R(q)$ 定义在具有重叠格的系统上, 关于 δ_n 的极限被 $\liminf_{\delta\to 0}$ 替换. 另外, Riedi (1995) 证明了 D_q 存在当且仅当 $\widetilde{\theta}_R(q)$ 是格正则的. 称 $\widetilde{\theta}_R(q)$ 是格正则的即用 $\lim_{\delta\to 0}$ 替换 $\liminf_{\delta\to 0}$ 极限存在. 这看起来并不是没有道理, 因为点中心构造和 Riedi 的构造都基于重叠盒子和一个光滑算子.

关于δ 的极限存在和本质

在 Rényi 维数 (全局平均) 和重分形谱 (局部行为) 中都考虑了用格和球作为覆盖的构造, 每个都具有趋于 0 的宽度. 在点中心情形, 用半径趋于 0 的球做覆盖, 假设极限对任意序列都存在. 在格的情形, 这样的极限不一定总存在, 所以需要考虑特殊序列.

Riedi (1995, 例 2.6.2) 证明了在 Cantor 测度的情形中, 若 $\widetilde{\theta}(q)$ 定义中的极限由 $\delta_n\to 0$ 换成 $\delta\to 0$, 则对 $q<1$, $\widetilde{\theta}(q)$ 不存在. 然而, 定理 5.5.4 说明对于带间隙的多项分布测度 $(q=2,3,\cdots)$, $\theta(q)$ 存在 (即当 $\delta\to 0$ 时), 这个证明基于 q 阶内点距离 (§2.4.5), 但是目前不知道这一思想能否扩充到负值的 q.

什么时候 $f(y)=\widetilde{f}(y)$?

在 $\theta(q)=\widetilde{\theta}(q)$, 并且重分形公式对格和点中心构造在弱意义下都成立时, $f(y)=\widetilde{f}(y)$.

注意到式 (5.11) 和 (5.12) 给出了点中心构造成立的关键条件. 方程 (5.11) 要求一定程度的光滑性和齐次性, 而式 (5.12) 要求当 $\delta\to 0$ 时幂等律的一致性. 假如对定理 5.5.2 中的多项分布测度成立且 $\theta(q)$ 满足要求, 那么 $f(y)=\dim_H F(y)$, 因此, $f(y)=\widetilde{f}(y)$. 注意到, 在这个例子中这是合理的, 因为这些测度在不同部分具有间隙, 导致 $f(y)$ 定义 (见式 (2.9)) 中的积分就是覆盖球的数目 (见注 2.5.3), 类似于 $\widetilde{f}(y)$ 定义 (式 (2.7)) 中覆盖盒子的数目.

第 6 章　倍增级联过程

6.1　引　　言

在第 4 和第 5 章中, 通过建立重分形体系来刻画空间 $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ 上概率测度 μ, 这里 $\mathcal{X}\subset\mathbb{R}^d$, $\mathcal{B}(\mathcal{X})$ 是 $\mathcal{X}$ 上的 Borel 集. 除多项分布测度外, 对测度 μ 的生成机制没有做任何假设. 本章假设测度 μ 是由级联过程产生的.

级联过程的研究起源于物理学, 特别是湍流. 这里能量大规模地进入系统, 空间和能量的数量都很巨大. 能量以不一致的方式扩散. 空间的某些部分可能会有漩涡并且有猛烈的行为, 但是其他部分可能很平静. 空间划分的越来越细, 能量慢慢地耗散, 最后变成热量. 这个一般的概念和第 3 章的多项分布测度很类似. 但是, 本章将作两个推广. 第一, 在每次划分过程中, 空间不一定等分. 第二, 能量耗散和概率分配可以随机进行.

可以把测度空间 $(\mathcal{X},\mathcal{B}(\mathcal{X}),\mu)$ 看做观察空间, 这里 $\mathcal{X}\subset\mathbb{R}^d$, μ 是所谓的重分形测度. 我们把它看做测度空间因为 μ 是一随机测度, 虽然要求其期望值为 1, 它将不再是一概率测度. 在其他情形, 像第 3 章的多项分布测度, μ 是一概率测度. 在前言中, 将建立能同时描叙这两种情形的一般框架. 级联过程涉及到在"粗糙"级别定义一个测度 μ_n, 然后 μ 定义成 μ_n 的极限.

测度 μ 的支撑由一迭代过程定义. 在过程的第 n 步, 需要一族不重叠的集合, 其并集包含测度 μ_n 的支撑. 在第 $n+1$ 步, 每个集合被细分, 如此下去. 例如, Cantor 测度的支撑含在 $[0,1]$ 中, 记作 $\mathcal{K}_0$. 第一级集合族为

$$\mathcal{K}_1=\left[0,\frac{1}{3}\right]\cup\left[\frac{2}{3},1\right];$$

第二级集合族为

$$\mathcal{K}_2=\left[0,\frac{1}{9}\right]\cup\left[\frac{2}{9},\frac{1}{3}\right]\cup\left[\frac{2}{3},\frac{7}{9}\right]\cup\left[\frac{8}{9},1\right];$$

等等. Cantor 集即为 $\mathcal{K}=\bigcap\limits_{n=0}^{\infty}\mathcal{K}_n$, 并且是 Cantor 测度的支撑. 这里需要一些记号来描叙更一般的框架.

类似第 3 章的多项分布测度, 一个简单的框架也可说明问题. 令 $\Omega=\{0,1,\cdots,b-1\}$, 这里 b 是固定的大于等于 2 的整数, 并且令 Ω^n 是由 Ω 中的数字组成长

为 n 的序列集合. 然后定义一个 Ω^n 和 $\mathcal{X}$ 非重叠子集之间的一一对应, 即假如 $\omega, \omega' \in \Omega^n$, 那么 $J_n(\omega)$ 和 $J_n(\omega')$ 不交. 因此, $\mathcal{K}_n$ 含有包含 μ 支撑的 b^n 个子集.

令 Ω^∞ 是由 Ω 中的数字组成无限序列集合. 定义 Ω^∞ 到 Ω^n 的投影为截取 $\omega \in \Omega^\infty$ 的前 n 项, 记为 $\omega|n$. 要求划分满足进一步的条件: 对任意的 n 和 $\omega \in \Omega^\infty$, $J_{n+1}(\omega|(n+1)) \subset J_n(\omega|n)$. 我们的目的是定义一个支撑为 $\mathcal{K} = \bigcap_{n=0}^{\infty} \mathcal{K}_n$ 的测度 μ. $\mathcal{K}_n$ 是测度 μ 支撑的一个覆盖, $\mathcal{K}_{n+1}$ 是一精细覆盖.

定义编码映射 X 如下:

$$X : \Omega^\infty \to \mathcal{X},$$

特别地,

$$X(\omega) = \bigcap_{n=1}^{\infty} J_n(\omega|n). \tag{6.1}$$

到此, 利用无限序列空间 Ω^∞ 描叙了一网状集列.

令 $\mathcal{S}_n$ 表示由 $\mathcal{K}_n$ 的所有子集 $J_n(x)$ 生成的子 σ-代数. 假设有一种办法, 对所有的 $A \in \mathcal{S}_n$, 定义一个测度 μ_n, 那么有一个测度空间 $(\mathcal{X}, \mathcal{S}_n, \mu_n)$. 将 μ 定义成 μ_n 的极限.

假如把 Ω^∞ 看做某个测度空间的一部分的话, 我们要小心. 回忆一下 Ω^n 中的元素和 $\mathcal{K}_n$ 的子集之间有一一对应关系, 因此, 由 $\mathcal{K}_n$ 的所有子集生成的子 σ-代数 $\mathcal{S}_n$ 和 Ω^n 生成的子 σ-代数 $\mathcal{F}_n$, 两者之间就有一一对应关系. 进一步, 假如 μ_n 已经定义好, 令 $\rho_n(E) = \mu_n(J_n(E))$, 这里 $E \in \mathcal{F}_n$. 因此, 两个测度空间 $(\Omega^\infty, \mathcal{F}_n, \rho_n)$ 和 $(\mathcal{X}, \mathcal{S}_n, \mu_n)$ 是同构的, 一个基于序列, 另一个基于 d-维欧氏空间. 下面的讨论总是利用方便的那一个.

下面希望在子集 $J_n(\omega)$ 上加一些更明确的结构, 并阐明怎样在这些子集上分配权重 (不一定是概率). 考虑以下映射:

$$W_i : \Omega^i \to (0, \infty) \tag{6.2}$$

和

$$T_i : \Omega^i \to (0, 1), \tag{6.3}$$

这里 $i = 1, \cdots, n$. W_i 描述了权重或质量在子集 $J_n(\omega)$ 上的分配, T_i 决定它们的大小. 令 μ_n 是第 n 步的质量分布, 并且假设质量分布和集合的大小可以分别表示如下:

$$\mu_n[J_n(\omega)] = \Pi_{i=1}^n W_i(\omega|i), \quad |J_n(\omega)| = \Pi_{i=1}^n T_i(\omega|i). \tag{6.4}$$

令 μ 为 $n \to \infty$ 时 μ_n 的极限质量分布.

为了刻画测度 μ 的重分形行为, 观察其局部行为. 在第 4 和 5 章, 考察如下形式的映射: $Y_n : \Omega^\infty \to \mathbb{R}$, 这里

$$Y_n(\omega) = \frac{\log \mu[J_n(\omega)]}{\log |J_n(\omega)|}, \tag{6.5}$$

其中当 $n \to \infty$ 时, $|J_n(\omega)| \to 0$. 注意 Y_n 不同于第 4 章的定义, 因为在本章, 对于所有 $\mathcal{K}_n$ 的子集, 分母 $\log |J_n(\omega)|$ 可能不一样.

关于级联这个词, 随机级联和决定性级联之间有些差别. 决定性在某种意义下表示映射 W_i 和 T_i 是明确定义的, 正如第 3 章的多项分布测度一样. 当 W_i 和 T_i 中的一个或者两个都是随机变量时, 就产生随机级联. 可能有 2×2 种情形, 可以根据以上意义下的确定性或随机过程来划分, 也可以根据确定性或随机过程来分配权重.

6.1.1 定义

假设形如式 (6.2) 和 (6.3) 的映射给定, 对所有的 $\omega \in \Omega^n, n = 1, 2, \cdots$ 给定子集 $J_n(\omega)$ 上的测度 μ_n 和 $J_n(\omega)$ 的大小. 还假设 μ_n 的极限存在, 记为 μ. 那么测度 μ 称为**倍增级联过程**.

μ_n 的局部行为记为 $Y_n^\ddagger(\omega)$, 表示如下:

$$Y_n^\ddagger(\omega) = \frac{\log \mu_n[J_n(\omega|n)]}{\log |J_n(\omega|n)|} = \frac{\displaystyle\sum_{i=1}^{n} \log W_i(\omega|i)}{\displaystyle\sum_{i=1}^{n} \log T_i(\omega|i)}, \quad \omega \in \Omega^\infty. \tag{6.6}$$

称之为**基本倍增级联过程**, 假如 $|J_n(\omega)| = \delta_n$, 这里 δ_n 对给定的 n 是常数, 并且当 $n \to \infty$ 时递减.

注意 $Y_n^\ddagger(\omega)$ 和 $Y_n(\omega)$ 是不同的, 这里 $Y_n^\ddagger(\omega)$ 是 μ_n 的局部行为, 而 $Y_n(\omega)$ 是 μ 的局部行为. 在 W_i 和 T_i 确定时, $Y_n^\ddagger(\omega) = Y_n(\omega)$. 假如权重的分配是随机的, 那么 $\mu_n[J_n(\omega)] \neq \mu[J_n(\omega)]$, 因此, $Y_n^\ddagger(\omega) \neq Y_n(\omega)$. 在多项分布测度 (第 3 章) 情形, 映射 T_i 是平凡的, 即 $T_i = b^{-1}$, 所以 $|J_n(\omega)| = b^{-n}$. 另外, 映射 W_i 也是明确定义好的, 因此, $Y_n^\ddagger(\omega) = Y_n(\omega)$.

在第 4 章, 全局平均、Rényi 维数、重分形谱和配分集分别用 $\widetilde{\theta}(q), \widetilde{D}_q, \widetilde{f}(y)$ 和 $\widetilde{F}(y)$ 表示. 同样, 在第 5 章, 它们分别用 $\theta(q), D_q, f(y)$ 和 $F(y)$ 表示. 在两章里, 所有覆盖的宽度都是一样的. 在本章, 将定义描述相似概念的函数, 并将其分别表示为 $\theta^*(q), D_q^*, f^*(y)$ 和 $F^*(y)$. 还将考虑随机测度. 这需要对 $\theta^*(q)$ 给出一个比 $\theta(q)$ 更一般的定义. 例如, 假如在 b-进制区间上构造一个随机测度 μ, 那么 $\widetilde{\theta}(q)$ 自然是随机的. 我们希望 $\theta^*(q)$ 是一个自然的累积母函数.

6.1.2 定义——全局平均

给定 n, q, $\theta^*(q)$ 定义为以下方程的解:

$$E\left[\frac{\mu_n^q\left[J_n(\omega)\right]}{|J_n(\omega)|\theta_n^*(q)}\right] = b^{-n}.$$

假设以上期望存在, 它是对概率为 b^{-n} 的 n-级基本区间取得. 等价的有以下类似于累积母函数的形式:

$$E\left[\exp\left(q\sum_{i=1}^{n}\log W_i(\omega|i) - \theta_n^*(q)\sum_{i=1}^{n}\log T_i(\omega|i)\right)\right] = b^{-n}.$$

那么, 如极限存在, $\theta(q) = \lim_{n\to\infty}\theta_n^*(q)$. 对 $q \neq 1$, 定义倍增级联 Rényi 维数为

$$D_q^* = \frac{\theta^*(q)}{q-1}.$$

即使在 $|J_n(\omega)| = \log\delta_n$ 时, 也要注意 $\widetilde{\theta}(q)$ 和 $\theta^*(q)$ 之间的差别. 当 μ 是一个随机测度时, $\widetilde{\theta}(q)$ 自然是随机的 (有时称为总体均值). 这点将在 §6.3 中进一步讨论. 然而, $\theta^*(q)$ 的解释类似于 $\widetilde{\theta}(q)$ 和 $\theta(q)$, 见例 3.3.2.

如果每一步的划分如第 4、5 章是正则的, 那么 $\widetilde{\theta}(q)$ 和 $\theta(q)$ 就有很重要的概率解释. 他们和重整累积生成函数有关, 正因为如此, 可以借助大偏差理论来描叙局部行为, 并验证全局平均和局部平均之间的 Legendre 变换关系. 在第 4 章, 在比较好的条件下, 测度 μ 的局部行为描叙为

$$\lim_{n\to\infty}\frac{\log\Pr\{Y_n \in A\}}{\log\delta_n} = -\inf_{y\in A} I(y),$$

这里 δ_n 是 n-级子集的直径, $I(y)$ 是通过 Legendre 变换与 $\widetilde{\theta}(q)$ 相关的量. 另外, 通过这种方式定义的重分形谱 $\widetilde{f}(y)$ 是和 $I(y)$(见式 (4.7)) 紧密相关的. 然而, 在目前讨论下, 每一步子集的直径不是常数 δ_n, 因此不能运用这些结果. 在级联情形下, 需要 $f^*(y)$ 的另一定义.

6.1.3 定义——重分形谱

倍增级联过程的重分形谱 $f^*(y)$ 定义如下:

$$f^*(y) = \inf_q\{qy - \theta^*(q)\}, \quad y \in \mathbb{R}.$$

在某种意义下, $f^*(y)$ 和由式 (2.7) 给出的 $\widetilde{f}(y)$ 类似. 当然, 这不完全正确. 回忆一下, 在第 4 章, $C_0(q)$ 是一致样本测度的重整累积母函数. 除了常数和符号差异, 它

和式 (4.6) 给出的 $\widetilde{\theta}(q)$ 相关. 然而, 按定义, $I_0(y)$ 是 $C_0(q)$ 的 Legendre 变换, 因此 $I_0(y)$ 和 $f^*(y)$ 类似. 注意到假如推论 4.3.5 中的条件满足, 那么由式 (2.7) 定义的 $\widetilde{f}(y)$ 满足 Legendre 变换关系. 目前, 由式 (2.7) 给出的 $\widetilde{f}(y)$ 不太重要, 因为对 $\mathcal{K}_n$ 的不同子集, $\log\delta_n$ 可能不相同.

本章剩下的内容主要讨论莫朗 (Moran) 和随机级联过程. §6.2 讨论莫朗级联过程. 这种情形比第 3 章的多项分布测度要复杂, 但是测度支撑和测度的分配是以确定的方式给出的. 在 §6.3, 回忆了一些和随机级联相关的结果. 这里研究的特殊性在于测度的分配是随机的, 但是子集的划分和第 3 章相同. 所以, 即使 $Y_n^{\ddagger}(\omega) \neq Y_n(\omega)$, Gartner-Ellis 定理仍然成立. 其他的级联过程在 §6.4 将作简单讨论.

本章大部分内容取自 Cawley 和 Mauldin (1992), Holley 和 Waymire, Gupta 和 Waymire (1990, 1993).

6.2 Moran 级联过程

级联过程起源于 Moran (1946 年) 的一个结果. 他证明了假如一个集合按某种方式细分, 极限集的 Hausdorff 维数可以由一个简单的关系表示. 遵守这一规则的集合是自相似的. 这一思想后来被 Hutchinson (1981) 发展. 自相似集的定义和他们的 Hausdorff 维数可以在附录 A, A.1 中找到.

Moran 分形集的构造和 Cantor 集类似. 不同的是比例不要求相同, 并且起始集合可以不是单位区间, 而是更一般的集合. Cawley 和 Mauldin (1992) 以及 Edgar 和 Mauldin (1992) 都讨论过 Moran 级联过程.

6.2.1 定义——Moran 分形集 (Moran, 1946; Hutchinson, 1981)

符号 $\bowtie$ 表示词的联接, 即假如 $\omega = (\omega_1, \cdots, \omega_n) \in \Omega^n$, 那么, $\omega \bowtie k$ 表示 $(\omega_1, \cdots, \omega_n, k) \in \Omega^{n+1}$. 每一个元素 $\omega \in \Omega = \{0, 1, \cdots, b-1\}$ 对应一个相似比例 t_ω, 满足 $0 < t_\omega < 1$. 假设 $\mathcal{K}_0$ 是起始集, 并且它是正则的, 即 $\mathcal{K}_0 = \overline{\mathrm{int}\mathcal{K}_0}$. 集合 $J_n(\omega), \omega \in \Omega^n$ 按如下方式递归定义. 不失一般性, 假设 $|\mathcal{K}_0| = 1$.

1) $J_1(0), \cdots, J_1(b-1)$ 是 $\mathcal{K}_0$ 不重叠的相似子集, 相似比分别为 $t_0, \cdots, t_{b-1}$ (不重叠就是满足开集条件, 见 Edgar 和 Mauldin, 1992, 606 页, 定义 A.1.1).

2) 假如 $J_n(\omega), \omega \in \Omega^n$ 已经定义, 那么 $J_{n+1}(\omega \bowtie 0), \cdots, J_{n+1}(\omega \bowtie b-1)$ 是 $J_n(\omega)$ 的不重叠的子集, 即对 $k \in \Omega, J_{n+1}(\omega \bowtie k)$ 是 $J_n(\omega)$ 的相似比为 t_k 的子集.

对 $n \geqslant 1, \mathcal{K}_n = \bigcup_{\omega \in \Omega^n} J_n(\omega)$. Moran 分形集定义为

$$\mathcal{K} = \bigcap_{n=0}^{\infty} \mathcal{K}_n.$$

6.2.2 定理

Moran 分形集的 Hausdorff 维数和盒维数相等, 另外,

$$\dim_H \mathcal{K} = \dim_B \mathcal{K} = s,$$

这里 s 是方程 $\sum\limits_{\omega\in\Omega} t_\omega^s = 1$ 的解.

证明 见定理 A.15. □

6.2.3 例子

假设图 6.1 左上角的圆表示起始集 $\mathcal{K}_0$. 令 $b=4$, $\mathcal{K}_1$ 由图 6.1 右上角 $(n=1)$ 的 4 块相似块 $J_1(0),\cdots,J_1(3)$ 组成. 它们相似比分别为 $t_0=\frac{1}{3}, t_1=\frac{1}{2}, t_2=\frac{1}{3}, t_3=\frac{1}{2}$. $\mathcal{K}_2$ 和 $\mathcal{K}_3$ 分别显示在图 6.1 的左下角和右下角. 由定理 6.2.2, 以上 Moran 分形集 $\mathcal{K}=\bigcap\limits_{n=0}^{\infty}\mathcal{K}_n$ 的 Hausdorff 维数为 $1.605525\cdots$.

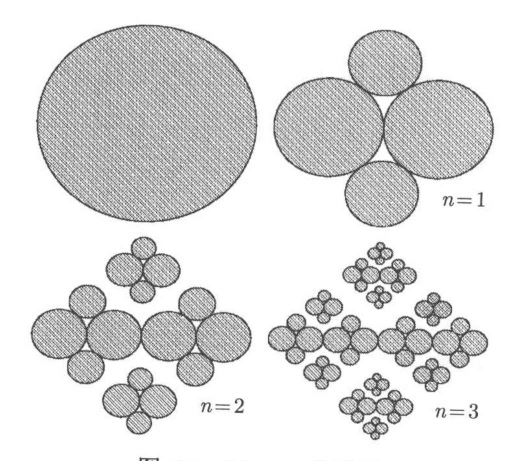


图 6.1 Moran 分形集

一个相似比为 $t_0=\frac{1}{3}, t_1=\frac{1}{2}, t_2=\frac{1}{3}$ 和 $t_3=\frac{1}{2}$ 的 Moran 集的例子, 其起始集是左上角的 K_0, K_1, K_2 和 K_3 如图所示. 其 Hausdorff 维数为 $1.605525\cdots$

6.2.4 Moran 级联测度

对每个 $\omega\in\Omega$, 赋予一非零概率 p_ω 满足 $\sum_{\omega\in\Omega} p_\omega = 1$ 和相似比 $0<t_\omega<1$. 那么由方程 (6.2) 和 (6.3) 定义的映射 W_i 和 T_i 可明确表示如下:

$$W_i(\omega)=p_{\omega_i},\quad T_i(\omega)=t_{\omega_i}.$$

因此

$$\mu_n[J_n(\omega)] = \Pi_{i=1}^n W_i(\omega|i) = p_{\omega_1}p_{\omega_2}\cdots p_{\omega_n},$$

并且

$$|J_n(\omega)| = \Pi_{i=1}^n T_i(\omega|i) = t_{\omega_1} t_{\omega_2} \cdots t_{\omega_n}.$$

测度 μ 是 μ_n 的极限, 并且其支撑是 Moran 分形集 $\mathcal{K}$.

注意与第 3 章的差别, 这里所有的 p_ω 都是非零的. 在本章的 Ω 上加上下标后两者也可以相容. 因为在第 3 章, $\Omega_0 = \{\omega : p_\omega > 0\}$. 为了简化记号, 不这样做. 另外, 在第 3 章的构造中, 零值的 p_ω 会使 μ 的支撑集产生间隙. 在本章是不必要的, 因为不同的相似比 (和比一定为 1) 也可以产生间隙.

由式 (6.6) 有

$$Y_n^{\ddagger}(\omega) = \frac{\log \mu_n[J_n(\omega|n)]}{\log |J_n(\omega|n)|} = \frac{\displaystyle\sum_{i=1}^n \log P_{\omega i}}{\displaystyle\sum_{i=1}^n \log t_{\omega i}}.$$

在 Moran 级联情形, $\mu[J_n(\omega)] = \mu_n[J_n(\omega)]$, 因此 $Y_n(\omega) = Y_n^{\ddagger}(\omega)$. 在第 4、5 章, 空间 $\mathcal{X}$ 按测度 μ 的局部幂定律划分, 即 $F(y) \subset \mathcal{X}$ 和 $\widetilde{F}(y) \subset \mathcal{X}$. 这里希望定义 $F^*(y)$ 使得 $F^*(y) \subset \mathcal{X}$.

6.2.5 定义——配分集

对于 $y > 0$, 配分集 $F^*(y)$ 定义为 $F^*(y) = X(G(y))$, 这里

$$G(y) = \{\omega \in \Omega^\infty : \lim_{n\to\infty} Y_n(\omega) = y\},$$

X 是式 (6.1) 定义的编码映射.

6.2.6 注记

回忆一下, 在第 4 章中, $B_{\delta_n}(x)$ 的定义为半开的. 因此, 对 $x \in \mathcal{X}$, $B_{\delta_n}(x)$ 是唯一的, 故集合

$$\widetilde{F}(y) = \left\{ x \in \mathcal{X} : \lim_{n\to\infty} \frac{\log \mu[B_{\delta_n}(x)]}{\log \delta_n} = y \right\}$$

有定义, 并且可以确定其局部行为.

Moran 分形集是具有正 μ-测度网状闭集合的交. 因此 $\mathcal{K}$ 是测度 μ 的支撑且 $\mathcal{K} \subset \mathcal{X}$. 另外, 因为 $\mathcal{K}_0$ 是闭的, 所以 $J_n(\omega)$ 也是闭的. 用码映射定义保证了 $F^*(y)$ 定义的合理性.

Moran 级联过程的函数 $\theta^*(q)$ 满足一个特殊简单的关系. Cawley 和 Mauldin (1992) 称 $\theta^*(q)$ 为"辅助"函数. 然而, 如前面所示, 它分别有相对于式 (2.2) 和 (2.5) 定义的 $\widetilde{\theta}(q)$ 和 $\theta(q)$ 的解释.

6.2.7 定理

在 Moran 级联过程中, $\theta^*(q), q \in \mathbb{R}$ 是方程

$$\sum_{k=0}^{b-1} p_k^q t_k^{-\theta^*(q)} = 1$$

的唯一解.

证明 由定义 6.1.2,

$$\sum_{\omega\in\Omega^n} \mu_n^q[J_n(\omega)]|J_n(\omega)|^{-\theta_n^*(q)} = \sum_{\omega\in\Omega^n} p_{\omega_1}^q \cdots p_{\omega_n}^q t_{\omega_1}^{-\theta_n^*(q)} \cdots t_{\omega_n}^{-\theta_n^*(q)}$$

$$= \left(\sum_{\omega\in\Omega} p_\omega^q t_\omega^{-\theta_n^*(q)}\right)^n.$$

以上和取遍 $\mathcal{K}_n$ 所有 b^n 个子集, 因此为 1. 因而所需结果得证. □

当 $q = 0$ 时, 由定理 6.2.2 有 $D_0^* = -\theta^*(0) = \dim_H \mathcal{K} = \dim_H \operatorname{supp}(\mu)$. 在多项分布测度情形, $t_i = b^{-1}$, 因此, $\theta^*(q) = \widetilde{\theta}(q)$.

6.2.8 定理——重分形谱

Moran 级联过程的重分形谱 $f^*(y), y \in \mathbb{R}$, 可按以下方式计算:

$$f^*(y_q^*) = qy_q^* - \theta^*(q),$$

这里 y_q^* 是 $\theta^*(q)$ 的导数, 即

$$y_q^* = \frac{d}{dq}\theta^*(q) = \frac{\sum_{k=0}^{b-1} p_k^q t_k^{-\theta^*(q)} \log p_k}{\sum_{k=0}^{b-1} p_k^q t_k^{-\theta^*(q)} \log t_k}.$$

证明 由定义 6.1.3, $f^*(y) = \inf_q\{qy - \theta^*(q)\}$. 设有两个关于 q 的函数, 一个是斜率为 y 的线性函数, 另一个是 $\theta^*(q)$, 见图 3.4. 对于给定的 q, 当两者具有相同的斜率, 即 $y = y_q^*$ 时, 达到下确界. □

函数 y_q^* 类似于第 4、5 章的 $y_m, m = q$.

6.2.9 定理——Hausdorff 维数 (Cawley 和 Mauldin, 1992, 定理 2.1)

令 $y > 0$, 那么 $\dim_H F^*(y) = f^*(y)$, 这里 $F^*(y)$ 是定义 6.2.5 中的配分集, $f^*(y)$ 是定理 6.2.8 中的重分形谱.

证明 证明分为两部分: 用 Vitali 覆盖定理证明 $\dim_H F^*(y) \leqslant f^*(y)$. 然后利用构造的特性来证明 $\dim_H F^*(y) \geqslant f^*(y)$. 在证明过程中, Cawley 和 Mauldin

(1992) 引进了空间 $(\Omega^\infty, \mathcal{F}_n)$ 上一个与概率 $(p_0 t_0^{-\theta^*(m)}, \cdots, p_{b-1} t_{b-1}^{-\theta^*(m)})$ 相关联的辅助测度, 记为 Π_m^∞. 由 $\theta^*(m)$ 的定义, 这些量之和为 1.

在第 4、5 章, 要求 $Y_n \to y_m$, ν_m^∞-几乎处处, 这里 ν_m^∞ 类似于 Π_m^∞. Cawley 和 Mauldin (1992) 借助于强大数定理和 Birkhoff 定理 (Walter 1982, 35 页). 对于 Π_m^∞-几乎处处的 $\omega \in \Omega^\infty$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log W_i(\omega|i) = -\sum_{k=0}^{b-1} p_k^m t_k^{-\theta^*(m)} \log p_k.$$

同样, 对于 Π_m^∞-几乎处处的 $\omega \in \Omega^\infty$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log T_i(\omega|i) = -\sum_{k=0}^{b-1} p_k^m t_k^{-\theta^*(m)} \log t_k.$$

因此, $\lim_{n\to\infty} Y_n(\omega) = y_m^*$, Π_m^∞-几乎处处. 因此, Ω^∞ 可以划分为 $\Pi_m^\infty[G(y_m)] = 1$. 故辅助概率 $(p_0 t_0^{-\theta^*(m)}, \cdots, p_{b-1} t_{b-1}^{-\theta^*(m)})$ 对应的乘积测度类似于第 4、5 章的样本测度. □

$F^*(y)$ 的定义的不足是: 需要知道收缩至 x 的集合序列 $J_n(\omega)$ 的构造, 才能确定何时 $x \in F^*(y)$. 对于不同类型的 Moran 构造, Cawley 和 Mauldin (1992) 证明了测度 μ 的局部行为可以用中心球表示.

6.2.10 定理 (Cawley 和 Mauldin, 1992)

考虑由不交映射生成的 Moran 级联过程 (即对于任意的 n, $\omega \neq \omega' \in \Omega^n$, $J_n(\omega) \cap J_n(\omega') = \varnothing$), 那么, $F^*(y) = F(y)$.

这是定理 5.5.1 和 5.5.2 的推广版本. 注意 $\mathcal{K}_n$ 的子集之间一定有间隙. 因为每个 $J_n(\omega)$ 都是闭集, 假如它们不相交, 则它们之间必有非零间距.

在一系列的论文中, Mandelbrot 与他的合作者讨论了带有不规则行为的倍增级联过程. 他们称之为“左侧精确自相似”过程, 其重分形谱在整个范围都是递增的 (Mandelbrot, 1990b; Mandelbrot et al., 1990; Riedi, Mandelbrot, 1995).

6.3 随 机 级 联

这种情形已经被 Holley 和 Waymire (1992) 及 Gupta 和 Waymire (1990,1993) 讨论过. 在本节, 这个过程起始集为 $\mathcal{K}_0 = [0,1]$, 然后将其等分成 b 份. 这些子区间将进一步细分. 区间 $J_n(\omega)$ 定义为

$$J_n(\omega) = \left[\sum_{j=1}^{n} \frac{\omega_j}{b^j}, \frac{1}{b^n} + \sum_{j=1}^{n} \frac{\omega_j}{b^j}\right].$$

在这种情形, 映射 T_i 是常值映射; 即对所有的 i 和 $\omega \in \Omega^i$, $T_i(\omega) = b^{-1}$. 因此, $J_n(\omega)$ 表示长为 b^{-n} 的特殊的子区间.

如同第 4 章, 有时用任意列举的区间代替 $J_n(\omega)$ 会更方便. 此时, 用 $B_{\delta_n}(k), k = 1, \cdots, b^n, \delta_n = b^{-n}$ 来表示. 包含 $x \in [0,1]$ 的盒子表示为 $B_{\delta_n}(k(x))$. 但是, 为了避免繁琐的记号, 当不会产生误会时, 记为 $B_{\delta_n}(x)$. 如第 4 章, $B_{\delta_n}(k), k = 1, \cdots, b^n$ 是右半开的, 但是 $J_n(\omega)$ 是闭区间. 然而, 正如多项分布测度, 所得测度在边界上的测度为 0, 所以不会影响结果.

权重将随机地分配到子区间上, 所以, 该测度不是概率测度. 就是说, 映射 $W_i(\omega)$ 不是一个随机变量.

6.3.1 随机测度的构造

按如下方式构造 $[0,1]$ 上一测度. 令

$$W_i(\omega) \equiv W(\omega_1, \cdots, \omega_i),$$

这里 $W_i(\omega), i = 1, 2, \cdots$ 是期望为 b^{-1} 的独立同分布的非负随机变量. 定义随机测度 μ_n 如下, 其密度 (Randon-Nikodym 导数) 为

$$\mu_n(dx) = b^n W_1(\omega) W_1(\omega) \cdots W_n(\omega) dx,$$

这里 $n \geqslant 1, x \in J_n(\omega)$. 因此

$$\begin{aligned}\mu_n[J_n(\omega)] &= \int_{J_n(\omega)} b^n W_1(\omega) W_2(\omega) \cdots W_n(\omega) dx \\ &= W_1(\omega) W_2(\omega) \cdots W_n(\omega),\end{aligned}$$

其极限测度记为 μ.

以下, 假设变量 W 和 $W_i(\omega)$ 具有相同的分布. 注意到 Gupta 和 Waymire (1993) 规定 $E[W] = 1$. 为了保证这个过程可以用式 (6.6) 来描述, 要求 $E[W] = b^{-1}$. 以下陈述一些相关结果, 主要来自 Holley 和 Waymire (1992).

6.3.2 定义 (Holley, Waymire 1992, 定义 2.1)

随机变量 W 称为强下有界的, 假如存在一个正数 a, 使得 $\Pr\{W > a\} = 1$. 同样地, W 称为强上有界的, 假如 $\Pr\{W < 1\} = 1$.

6.3.3 定理 (Kahane, Peyriere, 1976; Holley, Waymire, 1992)

假设 $W_i(\omega)$ 是和 W 具有相同的分布的变量. 令 $Z_\infty = \mu[0,1]$.

1) 假如 $E[W \log_b W] < 0$, 那么 $E[Z_\infty] > 0$, 相反, 条件 $E[Z_\infty] > 0$ 和 $E[Z_\infty] = 1$ 等价.

2) 令 $q>1$, 那么 Z_∞ 具有有限的 q 阶矩当且仅当 $q<q_{\text{crit}}=\sup\{q\geqslant 1: -\log_b E[W^q]>1\}$. 另外, 对 $q>0$, $E[Z_\infty^q]<\infty$ 当且仅当 1 是 W 的本性界 (即, $\|W\|_\infty<1$) 且 $\Pr\{W=1\}<1/b$.

3) 假设 $E[Z_\infty\log_b Z_\infty]<\infty$. 那么, μ-几乎处处被以下随机集合支撑

$$\operatorname{supp}(\mu)=\left\{x\in[0,1]:\lim_{n\to\infty}\frac{\log\mu[B_{\delta_n(x)}]}{\log\delta_n}=-bE[W\log_b W]\right\},$$

其 Hausdorff 维数为 $-bE[W\log_b W]$.

6.3.4 推论 (Holley, Waymire, 1992, 推论 2.5)

假如级联变量 W 是强上有界的, 那么 Z_∞ 具有所有的正阶矩. 假如级联变量 W 是强下有界的, 那么 Z_∞ 具有所有的负阶矩. 由于质量的分布是随机的, n 级区间的质量 $\mu_n[J_n(\omega)]$ 不一定和 $\mu[J_n(\omega)]$ 相等. 在随机级联文献中, μ_n 往往称为**赤裸级联测度**, 而 μ 往往称为**包装级联测度**. 以下命题阐述了这两个测度的分布.

6.3.5 命题 (Holley, Waymire, 1992, 命题 2.3)

设 $B_{\delta_n}(k), k=1,\cdots,b^n$ 是 $[0,1]$ 中 n-级 b-进制区间的一个列举. $\stackrel{d}{=}$ 表示具有相同的概率分布. 那么,

1)

$$\mu_m[B_{\delta_n}(k)]\stackrel{d}{=}Z_{m-n}^{(n)}(k)\mu_n[B_{\delta_n}(k)],$$

这里 $m\geqslant n, k=1,\cdots,b^n$, 并且 $Z_{m-n}^{(n)}(k)$ 和总质量 $\mu_{m-n}([0,1])$ 有相同分布, 与 $\mu_n[B_{\delta_n}(k)]$ 独立.

2)

$$\mu[B_{\delta_n}(k)]\stackrel{d}{=}Z_\infty^{(n)}(k)\mu_n[B_{\delta_n}(k)],$$

这里 $m\geqslant n, k=1,\cdots,b^n$, 并且 $Z_\infty^{(n)}(k)$ 和总质量 $\mu([0,1])$ 有相同分布, 与 $\mu_n[B_{\delta_n}(k)]$ 独立.

Holley 和 Waymire (1992) 用函数 $\chi_b(q)=\log_b E[W^q]+1$ 刻画了这一过程. 文献中 $\chi_b(q)$ 被称作 MKP(Mandelbrot-Kahane-Peyriere) 函数. 在随机级联过程中, 如定义 6.1.2 所示, $\chi_b(q)$ 是 $-\theta^*(q)$ 的逼近形式.

6.3.6 定理

设 μ 是随机级联过程的极限测度, 那么定义 6.1.2 中的 $\theta^*(q)$ 为

$$\theta^*(q)=-\log_b E[W^q]-1.$$

证明 此时, $\mu_n[J_n(\omega)] = W_1(\omega)\cdots W_n(\omega)$, 这里 $W_i(\omega)$ 是独立同分布的随机变量, 并且 $|J_n(\omega)| = b^{-n}$. 因此,

$$\begin{aligned} b^{-n} &= E\left[\mu_n^q[J_n(\omega)]|J_n(\omega)|^{-\theta_n^*(q)}\right] \\ &= (b^{-n})^{-\theta_n^*(q)}E[W_1^q(\omega)\cdots W_n^q(\omega)] \\ &= (b^{-n})^{-\theta_n^*(q)}E[W^q]^n, \end{aligned}$$

这里 W 和 W_i 具有相同分布. 所需结果得证. □

6.3.7 例——Log-Levy 生成子

一类级联生成子是由 $W = \exp(-Z)$ 给出的, 这里 Z 是一个 Levy 稳定随机变量. Lovejoy 和 Schertzer 及其合作者做了很多工作 (如, Schertzer, Lovejoy, 1989; Lovejoy, Schertzer, 1985, 1990), 他们提倡用这种过程来描叙降雨. Gupta 和 Waymire (1993, 式 (3.15)) 给出了这些过程的 $E[W^q]$:

$$\log E[W^q] = \begin{cases} -c\sigma^\alpha q^\alpha + \xi q, & 0 < \alpha < 1, \\ cq\log q - \xi q, & \alpha = 1, \\ c\sigma^\alpha q^\alpha - \xi q, & 1 < \alpha \leqslant 2, \end{cases}$$

这里 α 是特征指数, 当 $0 < \alpha < 2$ 时, $q > 0$, $\xi \in \mathbb{R}$ 为位置参数, $\sigma > 0$ 是标度参数, c 是正常数.

$0 < \alpha < 1$ 的情形可由 Feller (1971, § XIII.6, 定理 1) 得到, 因为 $E[W^q]$ 是 Z 的概率密度的 Laplace 变换. 其他推广形式见 Schertzer 和 Lovejoy (1987, 附录 C). 就像 Gupta 和 Waymire (1990, 29 页) 指出的, 所有的倍增生成子不在稳定吸引子范围内, 这些生成子通常描叙成“通用的” (Lovejoy, Schertzer, 1990). 另外, 还要注意这类生成子不满足定义 6.3.2 所述的强有界条件.

随机级联: 大偏差体系

我们已经注意到了赤裸测度和包装测度之间的差别以及 $\mu_n[J_n(\omega)] \neq \mu[J_n(\omega)]$. 因此, 这里可以讨论大偏差定理对 μ_n 及 μ 是否成立. 第 4, 5 章是从估计重整累积生成函数 $C(q)$ 开始, 通过 Legendre 变换推导熵函数 $I(y)$. 假如 μ_n 由随机变量 $\log W_i$ 的和定义, 那么很容易估计其全局和局部行为, 进而与 μ 的局部行为联系起来.

现在借助附录 B 的一些记号, 令 $\delta_n = b^{-n}$, $a_n = -\log\delta_n = n\log b$, 这里 $n = 1, 2, \cdots$. 另外, 定义映射

$$U_n : \Omega^\infty \to \mathbb{R},$$

这里

$$U_n(\omega) = -\log \mu_n[J_n(\omega|n)] = \sum_{i=1}^{n} \log W_i(\omega|i).$$

测度 μ_n 是随机的, 其构造方式见 §6.3.1. 进一步, 定义 $Y_n^{\ddagger}(\omega)$ 如下:

$$Y_n^{\ddagger}(\omega) = \frac{U_n(\omega)}{a_n} = \frac{\log \mu_n[J_n(\omega|n)]}{\log \delta_n} = \frac{-1}{n}\sum_{i=1}^{n} \log_b W_i(\omega|i).$$

6.3.8 重整累积生成函数

如第 4 章, 考虑所有 b-进制区间等长的情形. 对应的期望用 E_0 表示. 以下无下标的期望相对于概率分布 W 取的. 由方程 B.5,

$$\begin{aligned} C_0(q) &= \lim_{n\to\infty} \frac{1}{a_n} \log E_0[e^{qU_n(\omega)}] \\ &= \lim_{n\to\infty} \frac{1}{n} \log_b E_0[\mu_n^{-q}[J_n(\omega)]] \\ &= \lim_{n\to\infty} \frac{1}{n} \log_b E[W_1^{-q} \cdots W_n^{-q}] \\ &= \log_b E[W^{-q}]. \end{aligned}$$

注意到 $C_0(q) = -\theta^*(-q) - 1$. $\theta^*(q)$ 的导数记为 y_q^*, 即:

$$y_q^* = \frac{d}{dq}\theta^*(q) = \frac{-E[W^q \log_b W]}{E[W^q]}. \tag{6.7}$$

利用方程 B.6, 熵函数定义为

$$I_0(y) = \sup_q \{qy - C_0(q)\} = 1 - f^*(y).$$

6.3.9 定理——大偏差界

假设 $C_0(q)$ 满足广义假设 4.2.2, 那么

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log_b Pr\{-\varepsilon < Y_n^{\ddagger} - y \leqslant \varepsilon\} = -I_0(y), \quad y \in \operatorname{int}\mathcal{D}(I).$$

另外, $C_0(q) = \sup_y\{qy - I_0(y)\}$.

证明 幂定律的衰减方式和定理 4.2.4 相同. 有定理 B.3.17 得到 Legendre 变换. □

以上定理刻画了赤裸随机测度的局部行为. 同样可以描述装饰随机测度的局部行为.

6.3.10 推论

假设 $C_0(q)$ 满足扩张假设 4.2.2, 且 $E[W \log_b W] < 0$, 那么

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log_b Pr\{-\varepsilon < Y_n - y \leqslant \varepsilon\} = -I_0(y), \quad \text{对 } y \in \operatorname{int}\mathcal{D}(I).$$

证明 由命题 6.3.5,

$$\Pr\{-\varepsilon < Y_n - y \leqslant \varepsilon\} = \Pr\left\{y - \varepsilon < \frac{\log Z_\infty^n(k)}{\log \delta_n} + Y_n^\ddagger \leqslant y + \varepsilon\right\}.$$

假如 $E[W \log_b W] < 0$, 由定理 6.3.3 知, 测度 μ 是非退化的且 $E[Z_\infty] = 1$. 由定理 6.3.9 即得结果. □

Gupta 和 Waymire (1993, §3b) 利用大偏差的 Chernoff 定理直接得到以上结果.

随机级联: 总体和样本平均

由式 (2.2) 定义的函数 $\widetilde{\theta}(q)$ 刻画了当 $n \to \infty$ 时 $\sum_k' \mu^q[B_{\delta_n}(k)]$ 平均值的幂定律. μ 是一随机测度, 那么 $\widetilde{\theta}(q)$ 自然也是一随机测度, 它不会以随机的方式和 $\theta^*(q)$ 产生联系. 下面考虑 $\widetilde{\theta}(q)$ 的极限行为. 在随机级联的文献中, $\widetilde{\theta}(q)$ 被称为空间或样本平均, 而 $C_0(q)$ 和 $\theta^*(q)$ 被称为总体平均.

注意在 $\widetilde{\theta}(q)$ 的定义中分母用 $+\log \delta_n$. 在随机级联的文献中, 倾向于用 $-\log \delta_n$, 在非随机级联中用 $+\log \delta_n$. 这里要求这些定义在全书是相容的.

6.3.11 定理 (Holley, Waymire, 1992, 定理 2.7)

假设 W 是强上有界和下有界的, 且对给定的 q, $E[W^{2q}]/E[W^q]^2 < b$. 那么, 以概率 1, $\widetilde{\theta}(q) = \theta^*(q)$.

6.3.12 定理 (Holley, Waymire, 1992, 定理 2.8)

假设 W 是强上有界和下有界的, 且对所有的 q, $E[W^{2q}]/E[W^q]^2 < b$. 假如重分形谱 $\widetilde{f}(y)$ 存在, 那么, $f^*(y)$ 是 $\widetilde{f}(y)$ 的闭凸壳. 另外,

$$\theta^*(q) = \inf\{qy - \widetilde{f}(y)\}.$$

回忆一下, $C_0(y)$ 和 $I_0(y)$ 之间的 Legendre 变换关系两面都成立. 同时也注意到 §6.3.7 的 Log-Levy 生成子不满足强有界条件, 故以上结果不一定成立.

随机级联: Hausdorff 维数和盒维数

这里将介绍 Holley 和 Waymire (1992) 确定 Hausdorff 维数的方法. 第 4 章中虽然使用的是一族样本测度, 但是方法是类似的. 在随机级联中, 测度 μ 不是概率测度, 因此 Holley 和 Waymire (1992) 构造了一族对偶级联过程.

6.3.13 引理 (Holley, Waymire, 1992)

假设级联随机变量 W 是强上、下有界的, 且期望为 b^{-1}. 考虑使得 $\inf_q\{qy-\theta^*(q)\}=f^*(y)>0$ 的 y. 对 $y_q^*=y$ (方程 6.7) 的 y, 在 $q=\beta(y)$ 处有唯一解, 这里 $\beta(y)$ 是 y 的函数. 另外,

$$f^*(y)=\inf_q\{qy-\theta^*(q)\}=\beta(y)y-\theta^*(\beta(y)). \tag{6.8}$$

证明 同定理 6.2.8 的证明. □

6.3.14 定理 (Holley, Waymire, 1992, 定理 2.6)

假设级联随机变量 W 是强上、下有界的, 且期望为 b^{-1}. 还假设

$$E\left[\left(\frac{W}{\|W\|_\infty}\right)^{\beta(y)}\right]>\frac{1}{b},$$

这里 $\|W\|_\infty$ 表示 W 的本性界, $\beta(y)$ 由式 (6.8) 确定 (因此, 每个 y 满足 $f^*(y)>0$). 那么, $\dim_H \widetilde{F}(y)=\inf\{qy-\theta^*(q)\}=f^*(y)$, 这里

$$\widetilde{F}(y)=\left\{x\in[0,1]:\lim_{n\to\infty}\frac{\log\mu[B_{\delta_n}(x)]}{\log\delta_n}=y\right\}.$$

证明 证明概要如下. W 是期望为 1 的强有界的级联变量. 对满足 $\inf\{qy-\theta^*(q)\}=f^*(y)>0$ 的 y, 构造一个对偶的级联变量 W_β 使其具有如下分布:

$$W_\beta\overset{d}{=}\frac{W^\beta}{b^{\beta+1}E[W^\beta]},$$

这里 $y_\beta^*=y$(式 (6.7)). 即对所有的 $(\omega_1,\cdots,\omega_n)\in\Omega^n$, 用 $W_\beta(\omega_1,\cdots,\omega_n)$ 的样本值代替 $W(\omega_1,\cdots,\omega_n)$ 的样本值. 令 $\mu_{\infty,\beta}$ 表示对应的级联, $Z_{\infty,\beta}$ 表示总质量. 整个证明分为三部分.

1) 非平凡级联测度 $\mu_{\infty,\beta}$ 的存在性: 对所有的 $q>1$, $E[Z_{\infty,\beta}^q]<\infty$, $\mu_{\infty,\beta}$ 非平凡且以概率 1 为正的.

2) 划分空间.

(a) 可以证明, 以概率 1 有

$$\sup_{1\leqslant i\leqslant b^n}\frac{\log Z_\infty^{(n)}(i)}{n}\to 0,\quad 当\ n\to\infty,$$

这里 $Z_\infty^{(n)}(i), 1\leqslant i\leqslant b^n$ 是和 Z_∞ 同分布的独立同分布的随机变量; 并且

$$\sup_{1\leqslant i\leqslant b^n}\frac{\log Z_{\infty,\beta}^{(n)}(i)}{n}\to 0,\quad 当\ n\to\infty,$$

这里 $Z_{\infty,\beta}^{(n)}(i), 1 \leqslant i \leqslant b^n$ 是和 $Z_{\infty,\beta}$ 同分布的独立同分布的随机变量.

(b) 设 D_1 和 D_β 分别表示上述极限不存在的事件集. 因为 $D_1 \cup D_\beta$ 概率为 0, 可以仅考虑 $[0,1] \setminus (D_1 \cup D_\beta)$.

(c) 由定义, $x \in \widetilde{F}(y)$ 当且仅当

$$\lim_{n\to\infty} \frac{\log \mu[B_{\delta_n}(x)]}{\log \delta_n} = y. \tag{6.9}$$

由命题 6.3.5, $\mu[B_{\delta_n}(x)] \stackrel{d}{=} Z_\infty \mu_n[B_{\delta_n}(x)]$. 因此, 对 $x \in \widetilde{F}(y) \cap D_1^c \cap D_\beta^c$,

$$\begin{aligned}\lim_{n\to\infty} \frac{\log \mu[B_{\delta_n}(x)]}{\log b^{-n}} &\stackrel{d}{=} \lim_{n\to\infty} \frac{\log Z_\infty + \log \mu_n[B_{\delta_n}(x)]}{\log b^{-n}} \\ &\stackrel{d}{=} \lim_{n\to\infty} \frac{-1}{n}\left(\log_b Z_\infty - n + \sum_{i=1}^n \log_b W_i\right) \\ &\stackrel{d}{=} 1 - \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \log_b W_i.\end{aligned} \tag{6.10}$$

(d) 对 $x \in \widetilde{F}(y)$, 也估计测度 $\mu_{\infty,\beta}$ 的局部行为. 由命题 6.3.5, $\mu_{\infty,\beta}[B_{\delta_n}(x)] \stackrel{d}{=} Z_{\infty,\beta}\mu_{\infty,\beta}[B_{\delta_n}(x)]$. 因此, 对 $x \in \widetilde{F}(y) \cap D_1^c \cap D_\beta^c$,

$$\begin{aligned}\lim_{n\to\infty} \frac{\log \mu_{\infty,\beta}[B_{\delta_n}(x)]}{\log b^{-n}} &\stackrel{d}{=} \lim_{n\to\infty} \frac{\log Z_{\infty,\beta} + \log \mu_{n,\beta}[B_{\delta_n}(x)]}{\log b^{-n}} \\ &\stackrel{d}{=} \lim_{n\to\infty}\left(\beta + 1 - \frac{\log_b Z_{\infty,\beta}}{n} + \log_b E[W^\beta] - \frac{1}{n}\sum_{i=1}^n \log_b W_i^\beta\right) \\ &\stackrel{d}{=} \beta\left(1 - \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \log_b W_i\right) + \log_b E[W^\beta] + 1.\end{aligned}$$

然后, 由式 (6.9) 和 (6.10), 引理 6.3.13 和定理 6.3.6 有

$$\begin{aligned}\lim_{n\to\infty} \frac{\log \mu_{\infty,\beta}[B_{\delta_n}(x)]}{\log b^{-n}} &\stackrel{d}{=} \beta(y)y - \theta^*(\beta(y)) \\ &= \inf_q\{qy - \theta^*(q)\} \\ &= f^*(y).\end{aligned}$$

3) 利用 Billingsley 定理 4.5.6.

(a) 令

$$F_\beta(y) = \left\{x \in [0,1] : \lim_{n\to\infty} \frac{\log \mu_{\infty,\beta}[B_{\delta_n}(x)]}{\log b^{-n}} = f^*(y)\right\}.$$

因为以概率 1 有 $\mu_{\infty,\beta}F_\beta(y) = \mu_{\infty,\beta}([0,1])$, 那么

$$\dim_{\mu_{\infty,\beta}} F_\beta(y) = 1.$$

(b) 由构造, 有

$$\begin{aligned}\dim_H \widetilde{F}(y) &= \dim_H F_\beta(y)\\ &= f^*(y)\dim_{\mu_{\infty,\beta}} F_\beta(y)\\ &= f^*(y).\end{aligned}$$

□

6.3.15　命题 (Holley, Waymire, 1992)

假设 $\widetilde{F}(y)\neq\varnothing$, 那么 $\dim_B \widetilde{F}(y)=1$.

重分形谱 $f^*(y)$ 不一定表示整个范围上的几何维数. 假如定理 6.3.14 的条件满足, 那么当 $f^*(y)>0$ 时, $f^*(y)$ 可以看做 Hausdorff 维数. 但是, 经常会出现不得不考虑的使得 $f^*(y)$ 为负值的区域 (见下面的例 6.3.16). 这些值在文献中常被称为“负维数”. Mandelbrot (1989, 1990a, 1991) 的很多文献都讨论过这一现象. Mandelbrot 称使得 $f^*(y)>0$ 的值 y 为*显示的*, 那些使得 $f^*(y)<0$ 的值 y 为*隐藏的*. 另外, 那些使得 $y<0$ 且 $f^*(y)>-\infty$ 的值 y 称为*虚有的*. 这些情形可以从一下例子中显示.

6.3.16　例——对数正态分布

令随机权重定义为 $W=\exp(-Z)$, 这里 Z 是方差为 σ^2, 期望为 $\frac{1}{2}\sigma^2+\log b$ 正态随机变量 (见 §1.5.2; Kolmogorov, 1962; Oboukhov, 1962; Mandelbrot, 1989; Meneveau, Sreenivasan, 1991). 那么, W 是具有对数正态分布, 期望为 b^{-1} 的正随机变量.

这是例 6.3.7 中 Log-Lévy 生成子的一个特殊情况, 这里 $\xi=\frac{1}{2}\sigma^2+\log b, \alpha=2, c=\frac{1}{2}$. 由例 6.3.7,

$$\log E[W^q]=\frac{\sigma^2q^2}{2}-\frac{\sigma^2q}{2}-q\log b.$$

因此,

$$\theta^*(q)=-\psi q^2+(\psi+1)q-1,$$

这里 $\psi=\sigma^2/(2\log b)>0$. 因为 $E[W\log_b W]=(\psi-1)/b$, 由定理 6.3.3 得 $E[Z_\infty]=1$ 当且仅当 $\psi<1$; 即 $\mu([0,1])$ 具有期望值 1 当且仅当 $\psi<1$.

重分形谱 $f^*(y)$ 可以由 $\inf_q\{qy-\theta^*(q)\}$ 计算. 函数 $\theta^*(q)$ 是二次的, 因此给定 y, 求 $qy-\theta^*(q)$ 的极小值, 并设在 $q=q_{crit}$ 取得. 对 $\theta^*(q)$ 求导, 然后令其为 0 得到

$$y=-2\psi q_{\text{crit}}+\psi+1,$$

进而得到

$$q_{\text{crit}}=\beta(y)=\frac{1+\psi-y}{2\psi}.$$

回到 $f^*(y) = \beta(y)y - \theta^*(\beta(y))$ 并化简得到

$$f^*(y) = 1 - \frac{(y-\psi-1)^2}{4\psi}.$$

重分形谱 $f^*(y)$ 在图 6.2 中绘出. 注意到 $\psi > 0$, 因此, 当 $y \to \pm\infty$ 时, $f^*(y) \to -\infty$. 另外, 注意到对 $q \in \mathbb{R}$, $\theta^*(q)$ 有限, 且对所有的 $y \in \mathbb{R}$, $-\infty < f^*(y) \leqslant 1$. 这是图 6.2 重分形谱和图 3.2 中多项分布测度的重分形谱之间的重要差别. 在那种情况下, 函数 $\widetilde{f}(y)$ 的定义域和值域都是正的. 但是此时两者可以是负的.

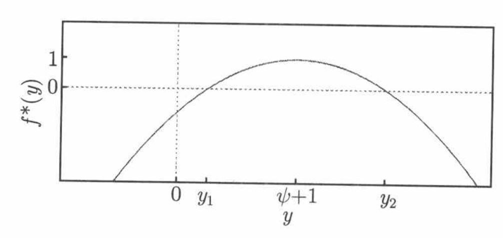


图 6.2 对数–正态级联的重分形谱

对数–正态随机级联过程的重分形谱. 当 $y_1 < y < y_2$ 时, $f^*(y)$ 是正的, 这里 $y_1 = \Psi + 1 - 2\sqrt{\Psi}$, $y_2 = \Psi + 1 + 2\sqrt{\Psi}$. 和图 3.2 中得多项分布测度相比, 那里对 $y < y_1$ 和 $y > y_2$, $\widetilde{f}(y) = -\infty$

以上谈到的两者之间的差别的原因是质量分配方式的不同. 在确定级联过程, 映射 W_i 的范围实际上是 $(0,1)$, 因此是概率分布. 进一步, 在每一步中, 倍增量 W_i 有固定的极大极小值. 这些值影响着 $f^*(y)$ 的可能取值 (见方程 (3.15) 和 (3.16)). 由构造, 概率在每一步是保留的, 只是在 $\mathcal{K}_n$ 的子集内重排而已.

在随机级联情形, 虽然 $E[W] = b^{-1}$, W_i 的允许值域为 $(0,\infty)$. 因此, "平均" 随机测度具有总质量 1. 假如 W 从长尾分布中取得, 那么可以取出非常大的值, 可能比 1 要大, 故 $Y_n^\ddagger < 0$. 这点解释了为什么 $f^*(y)$ 具有负值. 定理 6.3.9 给出了 $f^*(y) < 0$ 的解释. 可以从概率的角度解释函数 $I_0(y) = 1 - f^*(y)$ 是二次的且具有极小值 0. 对较小的值 y, $I_0(y)$ 刻画了 $Y_n^\ddagger$ 具有这样较小值的可能性, 或者, 等价的, 某个子集分配了很大的权重. 同样地, 对较大的值 y, $I_0(y)$ 刻画了某个子集分配了很小的权重的可能性. 推论 6.3.10 描叙了 Y_n 的这些特征, 即极限测度 μ 的局部行为.

6.4 其他级联过程

在 §6.2 和 §6.3 中级联构造都是基于确定的划分. 在紧集情形, 相似比例即使是随机的, Falconer (1986), Graf (1987) 及 Mauldin 和 Williams (1988) 证明了类似的结果仍然成立. Arbeiter (1991) 考虑了非紧情形.

随后, Falconer (1994) 考虑了统计自相似测度, 不但划分是随机的, 权重的分布也是随机的. 这点在 Olsen (1994) 也进一步讨论过.

Molchan (1995) 讨论过 Brown 的零点. 他用两个测度来分析. 抹去时间轴上距离原点 δ 的点来构造局部时间测度. 然后在剩下区间定义 Lebesgue 测度 $\lambda_\delta(dt)$. 局部时间测度是 $\delta \to 0$ 时正规化测度 $c\delta^{-1/2}\lambda_\delta(dt)$ 的极限. 在第二种情形, 他考虑 δ-丛数目的增长性. 在这个例子中, 划分和质量分配都带有随机性. 他建立了函数 $\theta^*(q)$ 和 $f^*(y)$, 并且证明它们之间由 Legendre 变换联系起来, 即 $\theta^*(q) = \inf_q\{qy - f^*(y)\}$.

6.4.1　$\theta^*(q)$ 的其他定义

$\theta^*(q)$ 有很多其他定义, 它们本质上和定义 6.1.2 类似. Halsey 等 (1986) 提出了 $\theta^*(q)$ 如下一个定义. 令

$$\mathcal{G}^{q,\xi}(F) = \lim_{\delta\to 0}\inf_{\{U_i\}}\left\{\sum_{i=1}^{\infty}\mu(U_i)^q|U_i|^{-\xi} : |U_i| \leqslant \delta, F \subset \bigcup_{i=1}^{\infty} U_i\right\},$$

这里下确界是对 F 的所有覆盖取的. 那么, 对任意的 q, 定义

$$\theta^*(q) = \inf\{\xi : \mathcal{G}^{q,\xi}(F) = 0\} = \sup\{\xi : \mathcal{G}^{q,\xi}(F) = \infty\}.$$

Pesin (1993, 542 页) 也讨论过这种情形, 他将其与一般维数建立了联系. 注意 $\dim_H(F) = -\theta^*(0)$. 这一定义对正则的覆盖有效, 不适用于随机分配测度.

许多作者提出了 $\theta^*(q)$ 的其他形式定义. Olsen (1994, 第二章) 给出了基于中心球覆盖的定义, 要求中心点位于集合 F 上. Olsen (1994) 也给出了基于填充维数 (见文献 Tricot, 1982; Falconer, 1990) 的定义, 那里 F 被中心在 F 上的填充球所覆盖.

第三部分

Rényi 维数的估计

第 7 章　q 阶点间距离和内在偏差

7.1　第三部分的引言

有许多理由解释为什么人们想刻画概率测度 μ 的重分形特征. 当概率测度包含尽可能多的不同阶数的奇点时, 诸如定位、扩展等用来刻画经典概率分布的措施是无用的. 若一个概率测度有多个不同阶数的奇点, 则有可能利用奇点的性质去刻画它. 重分形谱是一种描述 $\mathcal{X}$ 中包含给定阶数奇点的子集大小的一种方法. 然而, 用重分形谱处理数值是困难的, 从而人们通常估计 Rényi 维数, 然后计算产生重分形谱的 Legendre 变换 (第 2 章).

维数估计也描述测度 μ 在一个动力系统吸引集上非一致性的程度, 通常非线性方程 (非随机的) 很简单的集可以产生极其复杂的似乎是随机的行为. 维数估计已被用来确定一个时间序列是否已被一个典型低维的确定性过程或一个随机系统生成. 不过应当指出, 很多随机过程也有其标度服从幂律方式这样的性质, 一些概念和估算技术的讨论也可转移到这些情形.

伴随许多物理过程的问题, 特别是那种整个过程不可观察的问题, 比如说地震, 人们仅可记录由地震产生的地震波的特征, 由这些信息确定地震的其他特征. 气象系统中, 人们可以记录某些位置的测量信息, 比如降雨、风速以及云层. 结合卫星图像, 人们试图将它们拼凑成一个整体图片, 从而得到区域甚至全球气候. 在有些情形中, 仅一个标量时间序列被记录下来, 为了计算决定这个过程的基本方程的维数, 重新建立一个高维空间, 整个高维空间一般都显示出与这个不可观察系统具有相同的分形性质. 在第三部分中, 都假设 $\mathcal{X} \subset \mathbb{R}^d$ 是直接可观察的, μ 支撑在 $\mathcal{X}$ 的一个子集上. 这是因为嵌入的概念和重构为讨论增加了一层难度, 而这通常是不必要的, 嵌入和重构将在 §10.4.4 中讨论.

早期的研究涉及动力系统的分形维数, 维数可由宽为 δ 的 d-维盒子覆盖 d-维空间来估计, 覆盖的盒子数被计数, 这个过程对越来越小的 δ 值序列重复进行. 对以 $\log\delta$ 为单位的对数图, 在其直线部分所在的区域计算它的斜率. 这种方法涉及到基于格点的重分形, 特别是已在第 4 章中讨论过的 $\widetilde{D}_0$.

因为当 δ 充分小时, 大部分的盒子不被访问, 需要估计 $\widetilde{D}_0$ 的盒子数的计算量大. Grassberger 和 Procaccia(1983 a, b, c) 建议用一个替代的方法, 即淘汰需要计数访问的盒子, 大部分盒子被访问的次数为零, 他们主张人们应该通过估计点间距

离对概率分布的幂指数去充分认识维数. 不过, 这个维数是 D_2, 这是第 5 章中讨论过的内容, 在第三部分中的讨论也基于点间距离, 因此与第 5 章中重分形形式有关.

7.1.1 第三部分各章的概述

Grassberger 和 Procaccia(1983 a, b, c) 基于点间距离对的方法, 证明结果可推广到 q 阶点间距离 (如定理 2.4.5). q 阶点间距离可以用来计算高阶点中心 Rényi 维数, $q = 2, 3, \cdots$, 用 Y 表示 q 阶点间距离, 对 $q = 2, 3, \cdots$ 的情形估计 Rényi 维数, 主要涉及估计 Y 的概率分布函数 $F_Y(y)$ 的幂律指数.

点间距离的概率分布 $F_Y(y)$ 往往不是严格的幂律, 当概率测度支撑在分形集上时, 这种情况常常发生. 但是, 如果重对数图上的切线画在一个充分长的区间上, 则存在一个"平均的"幂律. 这可被认为是一种偏差的形式, 尽管它不是由抽样或其他方法上的缺陷造成的. 事实上它是过程的内在组成部分, 第 7 章讨论这种偏差的形式.

在第 8 章, 对 D_q, $q = 2, 3, \cdots$, 给出一种推广的 Grassberger-Procaccia 方法, 然而, 重点是 Hill 估计量以及其统计性质的修改后的版本. 这个估计量更自然地遵循极大似然估计的视角, 这对统计学家也许是一个前沿的挑战. 然而, 当使用 Grassberger-Procaccia 方法时, 有一个具有偏差的特殊问题不是非常明显, 这个偏差表现主要由第 7 章中讨论过的内在偏差所驱动.

许多真实数据仅仅有一部分令研究者有兴趣. 比如, 数据往往受到噪声的干扰. 此外, 这种噪声在时间或空间可能不均匀. 例如, 地震位置的记录在地震场所的高密度地区将更加准确, 准确度也随地震场所实际活跃的时间而变化, 所有从基本连续统中产生离散随机变量的数据被四舍五入. 另一个形式的偏差是由不观察整个过程, 在时间和空间上设置任意边界引起. 这会引起一个边界效应, 将这些归入偏差的外在形式, 在第 9 章中讨论它们的效应.

在第 10 章中介绍了一些维数估计的用途, 是应用数字和统计模型的技巧进行数据模拟, 而不是真实的数据. 运用我们部分了解性质的数据, 能更容易评价这个方法和解释结果.

第 11 章由一些使用真实数据研究的例子组成, 试图揭示维数图中有助于解释各种形式的偏差和数据缺失的那些部分, 这可能是分形真正可解释的内容.

7.1.2 符号的回顾

在 §2.4.1 中, $\theta(q)$ 定义为

$$\theta(q) = \lim_{\delta \to 0} \frac{\log \left[\int_{\mathcal{X}_\delta} \mu^{q-1}[S_\delta(x)]\mu(dx) \right]}{\log \delta}, \quad -\infty < q < \infty,$$

这里$\mathcal{X}_\delta=\{x\in\mathcal{X}:\mu[S_\delta(x)]>0\}$, 将用 $\theta(q)$ 表示 q 阶关联指数, 用 $\int\mu^{q-1}[S_\delta(x)]\mu(dx)$ 表示 q 阶关联积分, 点中心 Rényi 维数定义为

$$D_q=\frac{\theta(q)}{q-1},\tag{7.1}$$

这里 $q\neq 1, D_2=\theta(2)$ 常用来表示关联维数.

设 $X_1,X_2,\cdots,X_q$ 是由概率分布 μ 描述的独立随机变量的样本, 定义 Y (见定理 2.4.5) 为

$$Y=\max\{\|X_1-X_q\|,\|X_2-X_q\|,\cdots,\|X_{q-1}-X_q\|\}.$$

把 Y 叫做 q 阶内点距离, 注意 Y 由 q 决定, 为了避免符号笨拙, 没有用下标形式明确表示出来. 在计算中, 除非另有说明, $\|\,\|$ 总是表示 L^∞ 或最大模. 由定理 2.4.5, 对 $q=2,3,4,\cdots$, Y 的概率分布是

$$F_Y(y)=\Pr\{Y\leqslant y\}=\int\mu[S_y(x)]^{q-1}\mu(dx).\tag{7.2}$$

对 $q=2,3,4,\cdots$, 定理 2.4.5 的一个推论表明, 若下列极限存在, 则关联指数为

$$\theta(q)=\lim_{y\to 0}\frac{\log F_Y(y)}{\log y}.\tag{7.3}$$

□

在本章, 研究关联积分的特征, 特别研究支撑在一自相似集上的概率测度的情形. Rényi 维数是一内点距离逐渐变小的极限. 本章可看到 q 阶关联积分的内在性质, 这些性质可影响 Rényi 维数的估计.

7.2 边界效应

考虑随机样本内点距离 Y 的情形, 由 §2.4.3 知, 若概率测度 μ 可由概率密度函数表示, 那么 D_q 对 $q>0$ 而言是常数. 进一步, D_q 的值是 d, 即概率密度函数值域的维数. 由方程 (7.1) 和 (7.3), 对 $q=2,3,4,\cdots$, 有

$$D_q=\frac{1}{q-1}\lim_{y\to 0}\frac{\log F_Y(y)}{\log y}.$$

注意到 Rény 维数表示 $F_Y(y)$ 对 y 的无限小值的幂律指数, 同时这个幂律特征可不必扩展到较大的 y 值. 边界效应表示相对于由一严格幂律特征预测的较大点间距离的明显亏损, 随着内点距离的大小与区域的尺度可比, Y 值采样的机会减小.

在本节, 对几个简单情形分析边界效应, 在 §7.3 中可看到分形特征与这些边界效应的多样性有关.

7.2.1　L^∞ 模的分布

设 $q \geqslant 1$, X_i $(i = 1, 2, \cdots, q)$ 为 $\mathbb{R}^d$ 中独立随机变量, 即

$$X_i = (X_{i1}, X_{i2}, \cdots, X_{id})'.$$

进一步, 设 $X_{i1}, X_{i2}, \cdots, X_{id}$ 是独立的, 利用 L^∞ 模, Y 可表示为

$$Y = \max_{1\leqslant j\leqslant q-1} \|X_j - X_q\|_\infty = \max_{1\leqslant j\leqslant q-1} \max_{1\leqslant k\leqslant d} |X_{jk} - X_{qk}|.$$

假设对任意的 j, k, $|X_{jk} - X_{qk}|$ 与随机变量 U 具有相同的分布, 则

$$F_Y(y) = \Pr\{Y \leqslant y\} = [F_U(y)]^{d(q-1)}. \tag{7.4}$$

q 阶关联指数为

$$\theta(q) = d(q-1) \lim_{y\to 0} \frac{\log F_U(y)}{\log y}.$$

若 U 的密度存在, 则 $\theta(q) = d(q-1)$ 以及对 $q \geqslant 2$, $D_q = d$.　□

7.2.2　例子——Gauss 分布

考虑如下情形, $q = 2$, X_1 和 X_2 是 $\mathbb{R}^d$ 中多元正规随机变量, 利用 L^2 模, 差也是正规分布, 具有两倍方差, 从而

$$\begin{aligned}
F_Y(y) &= \int \mathbf{1}(\|x\|_2 \leqslant y) \frac{1}{\sqrt{2}} f_X\left(\frac{x}{\sqrt{2}}\right) dx \\
&= \int \mathbf{1}(\|\sqrt{2}x\|_2 \leqslant y) f_X(x)\, dx \\
&= \frac{|\Sigma^{-1}|^{1/2}}{(2\pi)^{d/2}} \int \mathbf{1}(\|\sqrt{2}x\|_2 \leqslant y) \exp\left(-\frac{1}{2}x'\Sigma^{-1}x\right) dx \\
&= \frac{|D^{-1}|^{1/2}}{(2\pi)^{d/2}} \int \mathbf{1}(\|\sqrt{2}z\|_2 \leqslant y) \exp\left(-\frac{1}{2}z'D^{-1}z\right) dz \\
&\qquad \text{这里} D = P'\Sigma P = \operatorname{diag}(\lambda_1, \cdots, \lambda_d) \\
&= \frac{1}{(2\pi)^{d/2}} \int \mathbf{1}(2w'Dw \leqslant y^2) \exp\left(-\frac{1}{2}w'w\right) dw \\
&\qquad \text{这里} w_i = z_i \lambda_i^{-1/2}, \quad i = 1, \cdots, d \\
&= \Pr\left\{2\sum_{i=1}^d \lambda_i W_i \leqslant y^2 \,\middle|\, W_i \text{是 i.i.d.} \mathcal{X}_1^2\right\}.
\end{aligned}$$

Davies(1980) 的程序能用来计算二次型概率.

当 $\Sigma = I$ 时, $F_Y(y)$ 可用具有 d 个自由度的 $\mathcal{X}^2$ 分布的观点表述, 如

$$F_Y(y) = \frac{2^{-d/2}}{\Gamma(d/2)} \int_0^{y^2/2} w^{d/2-1} \exp(-w/2)\, dw,$$

利用部分积分, 有

$$F_Y(y) = y^d \frac{\exp(-y^2/4)}{2^d} \sum_{k=0}^{\infty} \frac{1}{\Gamma(d/2+k+1)} \left(\frac{y^2}{4}\right)^k. \tag{7.5}$$

可以看到, $F_Y(y)$ 由两部分组成, 幂律部分具有指数为 d, 另一部分为非幂律部分, 非幂律部分单调递减, 当 $y \to 0$ 时, 具有极限 $[2^d\Gamma(d/2+1)]^{-1}$. 它有效地刻画一个边界效应, 该边界效应与一个纯幂律的发生相关, 且引起较大内点距离亏损.

注意, 通过在方程 (7.5) 中令 $d=1$ 并利用方程 (7.4), 也可计算得到一个 L^∞ 模下的关联积分的表达式, L^∞ 模与 L^2 模之间的比较如图 7.1 所示. □

Takens(1993, 245 页) 证明, 如果不同模仅仅差一个有界因子, 并且界限远离零, 则关联维数相同.

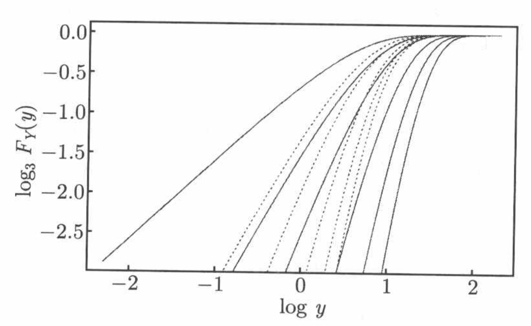


图 7.1 当 $q=2$ 时, 正规分布的关联积分

当 $q=2$ 时, 正规分布的关联积分在维数 $d=1$ (上面线),2, 3, 5, 7 和 9 (下面线) 的情形. 利用 L^2 模产生实线, 用 L^∞ 模产生虚线. 注意在 $d=1$ 的情形, 这些线相同, 当 $y \to 0$ 时, $\log y$ 关于 $\log F_Y(y)$ 的斜率趋于 d

7.2.3 例子——一致分布

设 $X_1, \cdots, X_q$ 是 $[0,1]^d$ 中独立连续一致随机变量, 具有独立分量, 则 q 阶内点距离具有概率分布 $F_Y(y) = y^{d(q-1)}(2-y)^{d(q-1)}$, 同时对 $q=2,3,\cdots$, 关联指数为

$$\theta(q) = \lim_{y\to 0} \frac{\log[y^{d(q-1)}(2-y)^{d(q-1)}]}{\log y} = d(q-1).$$

同样, 在例 7.2.2 中, Y 的概率函数具有幂律和非幂律部分, 非幂律部分描述边界效应. □

7.2.4 例子——“包围”度量

对一致随机变量的差, Theiler(1990) 通过利用一种“包围”度量证明边界效应消失. 即若 X_1 和 X_2 是在 $[0,1]$ 上一致独立同分布 (即 $d=1$), 则对 $q=2$,

$$Y = \|X_1 - X_2\| = \begin{cases} |X_1 - X_2|, & \text{若 } |X_1 - X_2| \leqslant \dfrac{1}{2} \\ 1 - |X_1 - X_2|, & \text{若 } |X_1 - X_2| > \dfrac{1}{2}. \end{cases}$$

在此情形下, $F_Y(y) = 2y$, $0 < y < \frac{1}{2}$, “非幂律部分”对 $y < \frac{1}{2}$ 均为常数, 这与在一个圆 $(d=1)$ 的圆周上或一个球 $(d=2)$ 的表面上一致分布点是相同的想法. 在这里, 两点间距离取最短弧长. □

7.3 边界的重数

在上一节, 边界效应被刻画成偏差, 因而从估计的观点来说是一件麻烦的事情, 这是不完全的情形. 事实上, 因为边界的重数以幂律方式发生, 故分形维数常常出现.

在例 7.2.3 中, 考虑单位区间上的一致分布, 其边界效应展示在图 7.2 中, 也可不一定在单位区间上考虑一致分布问题, 令 $\mathcal{K}_0$ 是一单位区间, $\mathcal{K}_1$ 是单位区间中间的三分之一去掉的部分, 一直下去, 就像 Cantor 集的构造. 这种情形与例 1.2.1 中描述的情形相同. 利用 $p_0 = \frac{1}{2}$, 使分配到每个子区间上的概率相等, 一直进行下去. 进一步注意到在例 3.1.8 中给出的 $G_n(y)$ 是在 $q = 2$ 时 $\mathcal{K}_n$ 上的点间距离的概率分布.

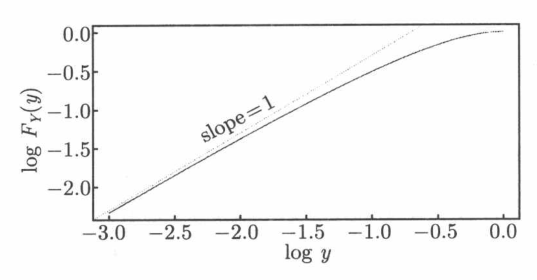


图 7.2　当 $q = 2$ 时, 一致分布的关联积分

对一致分布, $q = 2$, $d = 1$ 时的关联积分. 当 $y \to 0$ 时, 线的斜率趋于 1. 当 $y \to 1$ 时, 线中的弯曲由边界效应引起

对 $n = 0, \cdots, 7$, $\mathcal{K}_n$ 上一致分布的关联积分标在图 7.3 中, 最下面的实线是当一致分布支撑在 $\mathcal{K}_0$ 上的情形, 与图 7.2 相同, 从下数第二个实线是当测度支撑在 $\mathcal{K}_1$ 上的情形, 以此类推, 两条虚线仅供参考, 上面的虚线斜率是 $\log_3 2$, 下面虚线的斜率是 1. 注意 $p_0 = \frac{1}{2}$ 情形的 Cantor 测度的关联维数 D_2 是 $\log_3 2$.

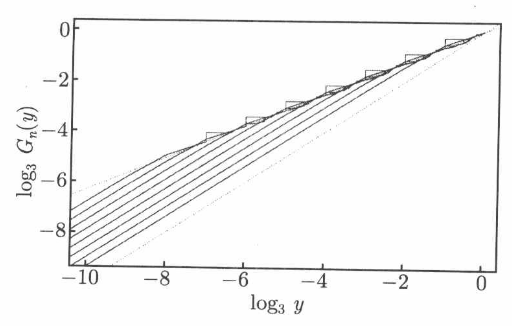


图 7.3 预 Cantor 测度的关联积分

对 $k=0,\cdots,7$, 预 Cantor 测度 $(p_0=0.5)$ 的关联积分 $(q=2)$. 下面虚线的斜率是 1, 是当点间距离较小时的渐近线, 上面虚线的斜率等于 D_2

考虑 K_1 的情形, 当点间距离小于 $\dfrac{1}{3}$, 本质上有两个分离的一致分布, 两者完全相同, 这里每个点对从相同的子区间开始, 因此对 y 取较小的值, 关联积分与当 $y\to 0$ 时一致分布有一个值为 1 的渐近斜率的线表现相同.

然而, 存在点间距离 y, 不可能有 $\dfrac{1}{3}<y<\dfrac{2}{3}$. 所以, 概率函数 $G_1(y)$ 在此区域将是平坦的, 一旦 $y\geqslant\dfrac{2}{3}$, 点间距离再次成为可能. 然而, 每个点必须来自每个不同的子区间. 当 $n=2$ 时, 在测度的支撑上, 添加另一个"洞". 在关联积分的图像中, 增加了另一个纽结, 每次去掉中间的 $\dfrac{1}{3}$, 另一个纽结就增加了. 当点间距离大于最小的"洞", 这条线的斜率大致是 D_2, 当点间距离小于最小的"洞", 其表现如同一致分布, 斜率趋于 1.

当 n 增加时, $G_n(y)$ 将收敛到 Cantor 测度的分布函数, 关联积分 $F_Y(y)$ 将很好地近似 $G_n(y)$, 如图 7.4.

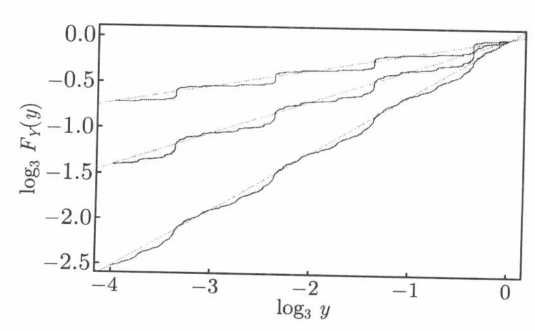


图 7.4 Cantor 测度的关联积分

对 $p_0=1-p_2, p_1=0$, 几种情形 Cantor 测度的关联积分, 最上面线是 $p_0=0.1$ 的情形. 中间线是 $p_0=0.2$ 的情形, 最下面线是 $p_0=0.5$ 的情形, 虚线具有斜率 $D_2=-\log_3(p_0^2+p_2^2)$

前面曾经提到挖掉子区间当做测度支撑里的“洞”. 图 7.4 中对数尺度的振动现象在文献中经常用来刻画缺项, 该词来自拉丁字 Lacuna, 表示“洞”的意思.

7.4　$F_Y(y)$ 的分解

本节中研究将 $F_Y(y)$ 分解成幂律和非幂律的分支, 使用 Theiler(1988) 的记号, 用 $\Phi(y)$ 表示非幂律分支.

7.4.1　命题

关联维数存在当且仅当关联积分能分解成

$$F_Y(y) = \Phi(y) y^{\zeta_q}, \tag{7.6}$$

这里 ζ_q 是一正的常数, $\Phi(y)$ 是一个正函数, 满足

$$\lim_{y \to 0} \frac{\log \Phi(y)}{\log y} = 0. \tag{7.7}$$

即当 $y \to 0$ 时, $|\log \Phi(y)| = o(|\log y|)$. 进一步, 给出分解 $\theta(q) = \zeta_q$.

证明　方程 (7.6) 和 (7.7) 蕴含 $\theta(q) = \zeta_q$, 反之若 $\theta(q)$ 存在, 则

$$\lim_{y \to 0} \frac{\log F_Y(y)}{\log y} = \theta(q).$$

因为

$$F_Y(y) = y^{\theta(q)} \frac{F_Y(y)}{y^{\theta(q)}},$$

所以

$$\lim_{y \to 0} \frac{\log(F_Y(y)/y^{\theta(q)})}{\log y} = 0,$$

从而有 $\Phi(y) = y^{-\theta(q)} F_Y(y)$. □

7.4.2　定义——自相似型行为

1. $F_Y(y)$ 称为具有标度参数 ζ_q 自相似的, 如果对所有 $s \in [0,1]$, 总有标度关系

$$F_Y(sy) = s^{\zeta_q} F_Y(y)$$

对所有满足 $y < y_{\max} = \inf\{F_Y(y) = 1\}$ 的 y 都成立.

2. $F_Y(y)$ 称为具有标度参数 ζ_q 的严格幂律行为, 如果对所有满足

$$y < y_{\max} = \inf\{F_Y(y) = 1\}$$

的 y 以及 $a = y_{\max}^{-\zeta_q}$ 都有

$$F_Y(y) = ay^{\zeta_q}.$$

3. 分布 $F_Y(y)$ 称为具有指数 ζ_q 的正则变化的下尾, 如果对所有 $s > 0$,

$$\lim_{y\to 0} \frac{F_Y(sy)}{F_Y(y)} = s^{\zeta_q}.$$

□

若 $F_Y(y)$ 是具有指数 ζ_q 的正则变化的下尾, 则

$$F_Y(sy) = s^{\zeta_q} F_Y(y) + o(F_Y(y)).$$

其中当 $y \to 0$ 时, $o(F_Y(y))/F_Y(y) \to 0$. 这样, 正则变化是自相似性的较弱形式.

7.4.3 命题

若 $F_Y(y)$ 是具有参数 ζ_q 的自相似, 则 $\theta(q)$ 存在且 $\theta(q) = \zeta_q$.

证明 由假设 $F_Y(y) = s^{-\zeta_q} F_Y(sy)$, 再由定义

$$\theta(q) = \lim_{y\to 0} \frac{\log F_Y(y)}{\log y} = \lim_{y\to 0} \frac{\log F_Y(sy)}{\log y}.$$

当 y 充分小, $F_Y(sy) = y^{\zeta_q} F_Y(s)$, 因此, 由上面方程,

$$\theta(q) = \lim_{y\to 0} \frac{\log[y^{\zeta_q} F_Y(s)]}{\log y} = \zeta_q.$$

□

7.4.4 命题

$F_Y(y)$ 是具有标度参数 ζ_q 的自相似当且仅当它具有参数 ζ_q 的严格幂律行为.

证明 严格幂律蕴涵自相似性可直接由定义得到. 反之, 若 $F_Y(y)$ 是自相似的, 则 $\theta(q)$ 存在且 $\theta(q) = \zeta_q$, $\theta(q)$ 的存在蕴涵存在一个形式分解 $F_Y(y) = \Phi(y)y^{\zeta_q}$. 给定自相似性, $\Phi(sy)(sy)^{\zeta_q} = s^{\zeta_q}\Phi(y)y^{\zeta_q}$, 即 $\Phi(sy) = \Phi(y)$ 对所有 0 到 1 之间的 s 和任意 y 成立, 所以 $\Phi(y) =$ 常数. □

Feller(1971, P276) 利用了函数在无穷远处缓慢变化的想法. 在本节中, 称函数 $\Phi(y)$ 在 0 处缓慢变化, 如果对所有 $s > 0$,

$$\lim_{y\to 0} \frac{\Phi(sy)}{\Phi(y)} = 1.$$

7.4.5 命题

已知 $\theta(q)$ 存在, 则 $F_Y(y)$ 有正则变化的下尾当且仅当 $\Phi(y)$ 在 0 处缓慢变化.

证明 因为 $\theta(q)$ 存在, 则存在 $F_Y(y)$ 的一个分解, 分解为幂律和非幂律部分, 另外, 还给出 $\Phi(y)$ 是一个在 0 处正则变化的函数

$$\lim_{y\to 0}\frac{\Phi(sy)(sy)^{\zeta_q}}{\Phi(y)y^{\zeta_q}}=s^{\zeta_q},$$

即 $\lim_{y\to 0}\dfrac{\Phi(sy)}{\Phi(y)}=1$. 反之亦然. □

在上面的讨论中, 已证明函数 $F_Y(y)$ 是自相似的等价于它是严格幂律的, 这样的情形限制太强. 然后, 考虑以下情形, 即对充分小于的 y, 它有类自相似的行为. 此外, 当点间距离是从概率分布中抽样, 且概率分布的测度支撑在一个分形集上, 此时在函数 $\Phi(y)$ 上强加的限制是不成立的. 我们需要一个比严格幂律行为弱的概念, 同时要求 $\theta(q)$ 仍然存在.

7.4.6 定义——缺项型行为

1. $F_Y(y)$ 称为具有标度参数 ζ_q 缺项的, 如果存在一个 0 到 1 之间的常数 c, 使得 $F_Y(cy)=c^{\zeta_q}F_Y(y)$ 对所有 $y<y_{\max}=\inf\{y:F_Y(y)=1\}$ 的 y 成立.

2. 分布 $F_Y(y)$ 称为具有指数 ζ_q 的在下尾处缺项的, 如果存在 0 到 1 之间的常数 c, 使得

$$\lim_{y\to 0}\frac{F_Y(cy)}{F_Y(y)}=c^{\zeta_q}.$$

□

注意缺项是自相似性较弱的形式.

7.4.7 命题

若 $F_Y(y)$ 是具有标度参数 ζ_q 缺项的, 则 θ_q 存在且 $\theta(q)=\zeta_q$. 进一步, $\Phi(y)$ 是一个在底为 c^{-1} 的对数标度上, 周期为 1 的有界周期函数.

证明 用与自相似性相同的方法得到 $\theta(q)=\zeta_q$, 由于缺项, 存在 0 到 1 之间的一个常数 c, 使得

$$\Phi(cy)(cy)^{\zeta_q}=c^{\zeta_q}y^{\zeta_q}\Phi(y).$$

即 $\Phi(cy)=\Phi(y)$, 则 $\Phi(y)$ 在底为 c^{-1} 的对数标度上是周期为 1 的周期函数. 给定方程 (7.7), 可选一区间, 即

$$I=(\log_{c^{-1}}y_0,1+\log_{c^{-1}}y_0),$$

这里 y_0 充分小使得对所有 $y\in I$, $|\log\Phi(y)|<|\log y|$, 即 $\log\Phi(y)$ 在 I 上有界. 由于 $\Phi(y)$ 是周期的, 则它一定是有界函数. □

7.4.8 例子

图 7.5 表示对 $p_0=0.5$ 的 Cantor 测度在 $q=2$ 时概率函数 $F_Y(y)$ 的非幂律分支 $\Phi(y)$. 事实上, 这个图像是 $\Phi(y)$ 的一个近似, 是由方程 (3.18) 给出的回复关系

经 11 次迭代得到的. 注意 $\Phi(y)$ 的振动行为有一个在对数标度下的常数周期和一个常数振幅, 它是关联积分刻画聚集性和在定义 7.4.6 意义下缺项的部分. □

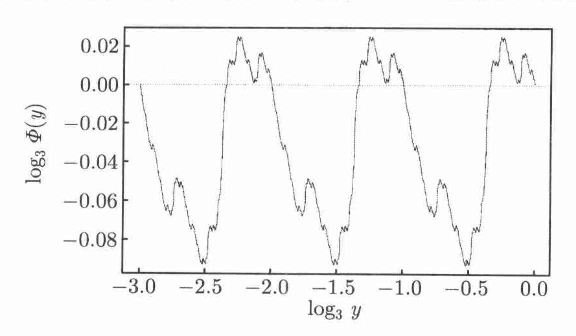


图 7.5 $p_0 = 0.5$ 时, Cantor 测度对应的 $\Phi(y)$

对 $p_0 = 1 - p_2 = 0.5$ 和 $p_1 = 0$ 的 Cantor 测度, 它是 $F_Y(y)$ 当 $q = 2$ 时的非幂律分支.

在本章讨论的 Cantor 测度的例子中, 函数 $\Phi(y)$ 在整个区域中是一个在底为 3 的对数标度下的周期函数, 从而 $F_Y(y)$ 在定义 7.4.6 意义下缺项, 表明对所有 $q = 2, 3, \cdots$ 成立. 不幸的是, 缺项对所有支撑在自相似集上的测度不能自动成立. 我们模拟的情形表明缺项依赖于"间隙"的位置. 在 §10.2 中, 将给出一些特殊的多项分布测度情形的例子, 其中 $\Phi(y)$ 不但包含一个类缺项的分支, 还包括一个边界效应分支, 因此, 在这类情形中 $F_Y(y)$ 仅在下尾中缺项 (见定义 7.4.6).

7.5 可 微 分 布

估计关联维数的 Grassberger-Procaccia 方法将在下一章讨论. 涉及到对某个 b, 描绘一个相对于 $\log_b y$ 的 $\log_b F_Y(y)$ 的估计图, 以及其斜率的估计. 令 $w = \log_b y$, 则此想法类似于相对于 w, 对 $\log_b F_Y(b^w)$ 取导数, 即假设 $F_Y(y)$ 是可微的

$$\frac{d}{dw} \log_b F_Y(b^w) = \theta(q) + b^w \frac{\Phi'(b^w)}{\Phi(b^w)}.$$

若 $\Phi(y)$ 单调递减, 上面的导数将小于 θ_q. 这种情形出现在例 7.2.2 和例 7.2.3 中, 表示相对于幂律行为较大点间距离的亏欠, Smith(1992 a, b) 的文章里包含了对 $q = 2$, $F_Y(y)$ 可微的情形的进一步分析.

第 8 章　点中心 Rényi 维数估计 ($q \geqslant 2$)

8.1　引　　言

在本章, 将探讨当 $q = 2, 3, \cdots$ 如何估计点中心 Rényi 维数 D_q, 以及它们的一些统计性质. 这些方法基于 q 阶内点差分法, 内点差分法在第 7 章已经讨论过. 在多数场合, 时间序列和动力系统只有一个标量过程是可以观察的. 从这个标量观察结果的序列中, 通常可以重新构造一个相位空间. 这一点将放在 §10.4.4 来讨论. 本章假设序列 $\{X_i\}$ 可以在 $\mathbb{R}^d$ 中得到, 测度 μ 由维数小于等于 d 的一个集合支撑.

8.1.1　符号回顾

令 $X_1, X_2, \cdots, X_q$ 为独立同分布的随机变量的样本, 且其概率分布为 μ. 定义 Y 为

$$Y = \max\{\|X_1 - X_q\|, \|X_2 - X_q\|, \cdots, \|X_{q-1} - X_q\|\}.$$

在计算中, 除非另外说明, $\|\ \|$ 指的是 L^∞ 意义下的最大模. 承接第 7 章, 不对 q 进行精确刻画, Y 的概率分布函数 $F_Y(y)$ 可由 (7.2) 式得出. 当 $q = 2, 3, \cdots$ 时, 相关指数 $\theta(q)$ 由 (7.3) 式可得; 从而由 (7.1), 得到 Rényi 维数 $D_q = \theta(q)/(q-1)$.

将 $\theta(q)$ 简记为 θ 以防符号过于复杂. θ 的估计值记为 Θ, 具体值记为 $\hat{\theta}$. 如果 $E[\Theta] = \theta$, 称 Θ 是无偏的; 如果对任意的 $\varepsilon > 0$, 有 $\lim_{N\to\infty} Pr\{|\Theta - \theta| > \varepsilon\} = 0$, 其中 N 为样本容量, 则称 Θ 是弱一致的. 弱一致性的充分条件是 Θ 无偏且当 $N \to \infty$ 时 $E[(\Theta - \theta)^2] \to 0$.

8.1.2　估计方法

最常用于估计 D_2 的方法是由 Grassberger 和 Procaccia(1983a, b, c) 给出的. 它是基于内点差分偶对的经验概率分布. 这一方法要求由 $\log F_Y(y)$ 确定的 $\log y$ 的图像, 在 $F_Y(y)$ 表现出幂行为的区域上, 能足够光滑为一条直线. 一个适当的区域需要被确定, 然后直线的斜率为相关维数的估计. 这一方法将在 §8.2 中陈述.

第二类估计是基于极大似然假设理论, 即在某种意义下 $F_Y(y)$ 具有幂律. 当前, 我们感兴趣的是在 y 取小数值情况下, $F_Y(y)$ 的幂律行为. 简单考虑特殊情况: $F_Y(y) = y^\theta$, $0 < y < 1$; 也就是对 y 的所有值, 给出 n 个独立内点距离的样本, log-似然方程为

$$L(\theta; y_1, \cdots, y_n) = n\log\theta + (\theta - 1)\sum_{i=1}^{n}\log y_i,$$

那么 θ 的极大似然估计为

$$\hat{\theta} = \frac{n}{-\sum_{i=1}^{n}\log y_i}.$$

注意到指数分布的关系, 随机变量 $W = -\log Y$ 的密度为 $\theta\exp(-\theta\omega)$.

更一般的情况是, 假设对 y 足够小的值, $y \leqslant \varepsilon$, $F_Y(y) = ay^\theta$, 其中 a 是正的常数. 假设分布在下尾部具有幂律, 而不对分布函数假设任何全局参数的形式. 因此这一估计方法使得 Rényi 维数与极值理论的方法相似; 详细参看, Embrechts 等.(1997). 内点距离的样本集为 $y_1, \cdots, y_n$, 考虑两种不同的调节机制:

1. 对样本 $y_1, \cdots, y_n$, 考虑那些值小于 ε 的 $N(\varepsilon)$. 这里 $N(\varepsilon)$ 是二项式随机变量, 参数为 n 和 $a\varepsilon^\theta$. 应用这一方法, 可以得到 Takens(1985) 给出的估计量, 这一估计称为 Takens 估计. 这些将在 §8.3 深入讨论.

2. 把样本调整为顺序统计量 $y_{(1)} \leqslant y_{(2)} \leqslant \cdots y_{(m)} \leqslant \cdots \leqslant y_{(n)}$. 考虑 $y_{(m)} \leqslant \varepsilon$. 由 Hill(1975) 的方法, 对极端事件, 估计在厚尾分布下的幂指数, 该估计称为 Hill 估计. 这些将在 §8.4 深入讨论.

许多可观察到的过程与时间演化是相关的. 因此, 选取内点差分的样本是相互独立的这一假设存在问题, 尤其在 Grassberger-Procaccia 算法中, 这一算法要求所有可能的差分通常都包括在计算中. 物理学家通常将时间序列稀疏化, 取每组第 m 个观察值, 其中 m 足够大. 作为一种替代, 要对内点距离随机取样, 同时要提取很多自举样本. 这一方法会在 §8.5 进行讨论. 在最后一部分, 我们会分析不同方法的优劣.

8.2 推广的 Grassberger-Procaccia 运算法则

8.2.1 $q = 2$ 的情形

Grassberger 和 Procaccia(1983a) 考虑了当 $q = 2$ 的情形. 给定一个 $\mathbb{R}^d$ 中的矢量随机向量的有限序列 $X_1, X_2, \cdots, X_N$, 他们的方法是计算所有可能的内点距离, 并用这些建立一个经验分布函数作为 $F_Y(y)$ 的估计. 即对 $q = 2$,

$$\hat{F}_Y^{GP}(y, N) = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbf{1}(\|X_i - X_j\| \leqslant y). \tag{8.1}$$

其中如果 A 发生, $\mathbf{1}(A)$ 是 1, 否则为 0. 下面将根据 $\log\hat{F}_Y^{GP}(y, N)$ 的样本值, 对 $\log y$ 画图, 并且把一些适当区域上图像的斜率作为相关维数 $D_2 = \theta(2)$ 的估计.

8.2.2　$q = 2, 3, \cdots$ 的情形

给定一个 $\mathbb{R}^d$ 中的随机向量的有限序列 $X_1, X_2, \cdots, X_N$, 定义 $\hat{F}_Y^{GP}(y, N)$ 为

$$\hat{F}_Y^{GP}(y, N) = \frac{1}{m}\sum_{k=1}^{m} \mathbf{1}(Y_k \leqslant y), \tag{8.2}$$

其中 $Y_1, \cdots, Y_m$ 为给定的容量为 N 的样本中第 q 阶差分的所有可能排列, 第 k 个记为

$$Y_k = \max\{\|X_{k_1} - X_{k_q}\|, \|X_{k_2} - X_{k_q}\|, \cdots, \|X_{k_{q-1}} - X_{k_q}\|\}.$$

当 $q > 2$ 时, 从 $X_1, X_2, \cdots, X_N$ 中不放回的取样记为 X_{k_q}, 再从剩余中的有放回取样记为 $X_{k_i}, i = 1, \cdots, q-1$. 因此, 在第二步有 N^{q-1} 种可能, 在第一步有 N 种可能. 这样 $m = N(N-1)^{q-1}$. 函数 $\hat{F}_Y^{GP}(y, N)$ 为 (7.2) 式中 $F_Y(y)$ 的估计.

8.2.3　指数 ν

相关维数 ($q = 2$) 的定义常常不同 (Theiler,1986, 方程 4), 其中

$$\nu = \lim_{y\to 0}\lim_{N\to\infty} \frac{\log \hat{F}_Y^{GP}(y, N)}{\log y}.$$

如果 $q = 2$, 相关指数 $\theta(2)$ 与相应的 Rényi 维数 D_2 相同. 称 D_2 为相关维数. ν 的极限形式与估计量很相似; 在定义 (7.3) 下, 如果一个适当的大数定律成立, ν 就等于 D_2. 当大数定律不成立时, ν 的解释是有问题的.

8.2.4　斜率的确定

直线的斜率需要被估计的时候, 间距的确定是非平凡的. 理想地, 想要挑选一个区间 (y_1, y_2), 其中 $y_{\min} \leqslant y_1 < y_2 \leqslant y_{\max}$, 且曲线上的弦与 $\log F_Y(y)$ 有相同长期依赖的斜率, 则

$$\hat{\theta}(q) = 估计斜率 = \frac{\log \hat{F}_Y^{GP}(y_2, N) - \log \hat{F}_Y^{GP}(y_1, N)}{\log(y_2) - \log(y_1)}.$$

我们用到的方法取决于 $\Phi(y)$ 的行为. $\Phi(y)$ 的特征开始并不知道. 当 $\Phi(y)$ 单调减时, 存在一个较大内点距离的空缺, 这一空缺由边界效应产生. Nerenberg 和 Essex(1990), Essex 和 Nerenberg(1991) 等指出, 边界效应和数据集的有限性影响了 $\hat{F}_Y^{GP}(y, N)$ 的值. 通过把 y 的较小值作为目标, 边界影响就几乎没有了. 通过较小的二进制计数会导致更大的统计误差. 由于 y 的值较小, 数据中的噪声也会带来问题. 边界带来的偏差和数据中的噪声会在第 9 章中深入探讨.

在不是全部由边界效应引起的的缺项的情况下, 不存在取小 y 样本的必要性, 但是需要画出一条弦, 使其不会被周期性误导. 更多内容参见 Eckmann 和 Ruelle (1992) 及 Ruelle(1990).

8.2.5 遍历过程的相容性

§1.3 提到过动力系统. Cutler(1991)、Denker 和 Keller(1986)、Pesin(1993) 都讨论过相容性. 一般情况下, 能描绘出一个过程 (X, T_ξ, μ), 其中映射 $T_\xi : \chi \to \chi$, μ 是 $B(\mathcal{X})$ 上的一个概率测度, 且映射是保测变换, 即 $\mu = T_\xi^{-1}\mu$; 等价地, μ 对 T_ξ 是不变的. 假定这样的系统是遍历的, 使得时间平均关于 μ 对几乎所有的 $x(t_0) \in \chi$ 收敛, 其中 $x(t_0)$ 为序列中的初始值. 这样的过程常常需要被混合, 即对所有的 Borel 集 $A, B \subset \chi$,

$$\lim_{n\to\infty} \mu[A \cap S^{-n}(B)] = \mu(A)\mu(B).$$

Pesin(1993) 已经证明: 如果序列是由像 $x(t_n) = T_\xi^n(x(t_0))$, $n = 1, 2, \cdots$ 生成的, 而且 μ 是遍历的, 则关于 μ 几乎所有 $x(t_0) \in \chi$, 有

$$\lim_{N\to\infty} \hat{F}_Y^{GP}(y, N) = F_Y(y).$$

Denker 和 Keller(1986) 证明了, 如果 (χ, T_ξ, μ) 是一个平稳遍历动力系统, 且满足某些混合条件, 则对 $y > 0$, $\hat{F}_Y^{GP}(y, N)$ 和 $\log \hat{F}_Y^{GP}(y, N)$ 分别依概率收敛于 $F_Y(y)$ 和 $\log F_Y(y)$. 他们也证明了当 $N \to \infty$ 时, $\hat{F}_Y^{GP}(y, N)$ 渐近于正态分布. Cutler(1991) 讨论了 $\log \hat{F}_Y^{GP}(y, N)$ 退化到 $\log y$ 产生的最小均方估计的问题和性质. Mikosch 和 Wang 证明了: 如果 $\{X_i\}$ 是一个平稳的遍历序列, 并且其 d 维随机向量有连续分布 μ, 且 μ 和 $\{X_i\}$ 具有同样的不变概率测度, 那么当 $N \to \infty$ 时, 有 $\hat{F}_Y^{GP}(y, N) \stackrel{\text{a.s.}}{\to} F_Y(y)$.

8.3 Takens 估 计

首先, 假设 y_i 独立, 并且 $F_Y(y) = ay^\theta$, $a > 0$. 根据 $y_1, \cdots, y_n$ 样本点的取值, 限定 $N(\varepsilon)$ 的值小于 ε. 那么可以得出条件分布函数

$$F_{Y|Y<\varepsilon}(y) = \frac{F_Y(y)}{F_Y(\varepsilon)} = \left(\frac{y}{\varepsilon}\right)^\theta.$$

对数似然方程可以写为

$$L(\theta; y_1, \cdots, y_{N(\varepsilon)}) = N(\varepsilon)\log\theta - N(\varepsilon)\log\varepsilon + (\theta - 1)\sum_{i=1}^{N(\theta)} \log\left(\frac{y_i}{\varepsilon}\right).$$

相应的 θ 的极大似然估计为

$$\Theta_\varepsilon = \frac{N(\varepsilon)}{-\sum_{i=1}^{N(\varepsilon)} \log\left(\frac{Y_i}{\varepsilon}\right)}.$$

这个估计是由 Takens(1985) 给出的, 其成立条件为 $Y_i < \varepsilon$. 当 $F_Y(y) = \Phi(y)y^\theta$ 时, 我们有兴趣确定 Θ_ε 的特性.

8.3.1　命题 (Theiler, 1988)

如果 $\lim_{\varepsilon\to 0}\gamma(\varepsilon) = 0$, $\dfrac{1}{\Theta_\varepsilon}$ 是 $\dfrac{1}{\theta}$ 的渐近无偏估计, 其中 $\gamma(\varepsilon) = \int_0^\infty \left[\dfrac{\Phi(\varepsilon e^{-\omega})}{\Phi(\varepsilon)} - 1\right]\theta e^{-\theta\omega}d\omega$, 并且当 $y \to 0$ 时, $\Phi(y) \to$ 常数. 特别地,

$$E\left[\frac{1}{\Theta_\varepsilon}|Y \leqslant \varepsilon\right] = \frac{1}{\theta}(1+\gamma(\varepsilon)). \tag{8.3}$$

证明　设 $W_i = -\log\left(\dfrac{Y_i}{\varepsilon}\right)$, 那么 $W_i > 0$ 当且仅当 $Y_i < \varepsilon$.

$$\Pr\{W_i > \omega|W_i > 0\} = \frac{\Pr\{W_i > \omega\}}{\Pr\{W_i > 0\}} = \frac{F_Y(\varepsilon e^{-\omega})}{F_Y(\varepsilon)} = \frac{\Phi(\varepsilon e^{-\omega})}{\Phi(\varepsilon)}e^{-\theta\omega}.$$

由分部积分, 当 $F_W(0) = 0$ 时, $E[W] = \int \omega f_W(\omega)d\omega = \int [1 - F_W(\omega)]d\omega$. 通过直接积分有

$$\begin{aligned} E\left[\frac{1}{\Theta_\varepsilon}|Y \leqslant \varepsilon\right] &= E[W_i|W_i > 0] \\ &= \frac{1}{\Phi(\varepsilon)}\int_0^\infty \Phi(\varepsilon e^{-\omega})e^{\theta\omega}d\omega \\ &= \frac{1}{\theta}\left(1 + \int_0^\infty \left[\frac{\Phi(\varepsilon e^{\omega})}{\Phi(\varepsilon)-1}\right]\theta e^{-\theta\omega}d\omega\right). \end{aligned}$$

□

Theiler(1988,1990) 使用了相关积分的一个更清晰的关系; 即通过用 $y = \varepsilon e^{-\omega}$ 替换上式得出

$$E\left[\frac{1}{\Theta_\varepsilon}\middle|Y \leqslant \varepsilon\right] = \frac{1}{F_Y(\varepsilon)}\int_0^\varepsilon \frac{F_Y(y)}{y}dy.$$

由分部积分得出 $\gamma(\varepsilon)$ 的级数展式

$$\gamma(\varepsilon) = \frac{1}{\Phi(\varepsilon)}\sum_{k=1}(-\varepsilon)^k\frac{\Gamma(\theta+1)}{\Gamma(\theta+k+1)}[\Phi^k(\varepsilon) - \lim_{\omega\to\infty}\Phi^k(\varepsilon e^{-\omega})e^{-(\theta+k)\omega}].$$

如果 $\Phi(y)$ 的第 k 阶导数 $\Phi^{(k)}(y)$ 对任意小的 y 是有界的, 并且 $\lim_{y\to 0}\Phi^{(k)}(\varepsilon y)y^{\theta+k} \to 0$, 那么 $\gamma(\varepsilon) = \dfrac{1}{\Phi(\varepsilon)}\sum_{k=1}(-1)^k\varepsilon^k\dfrac{\Gamma(\theta+1)}{\theta+k+1}\Phi^{(k)}(\varepsilon)$. 如果导数为零, 则和是有限的.

我们感兴趣的是 $E[\Theta_\varepsilon|Y \leqslant \varepsilon]$, 而不是 $E\left[\left.\frac{1}{\Theta_\varepsilon}\right|Y \leqslant \varepsilon\right]$. 通过下面的例子可以看到它们的差异.

8.3.2 例子

假设独立同分布指数随机变量序列 $W_1, \cdots, W_m$ 满足 $E[W_1] = \frac{1}{\theta}$, 那么 θ 的极大似然估计为

$$\Theta = \frac{m}{\sum_{i=1}^{m} W_i}.$$

显然, $E[1/\Theta] = 1/\theta$, 而 $E[\Theta] = \frac{m}{m-1}\theta$. □

8.3.3 命题

已知 $F_Y(y) = \Phi(y)y^\theta$ 及对所有的 $k > 1$ 都满足 $E[(\Theta_\varepsilon - \theta)^k|Y < \varepsilon] < \infty$, 那么

$$E[\Theta_\varepsilon|Y < \varepsilon] = \theta(1 - \gamma(\varepsilon)) + \frac{E[(\Theta_\varepsilon - \theta)^2|Y < \varepsilon]}{\theta} - \cdots.$$

证明 考虑 $1/\Theta_\varepsilon$ 关于 $1/\theta$ 的泰勒展式, 取期望可得

$$E\left[\left.\frac{1}{\Theta_\varepsilon}\right|Y < \varepsilon\right] = \frac{1}{\theta} - \frac{E[(\Theta_\varepsilon - \theta)|Y < \varepsilon]}{\theta^2} + \frac{E[(\Theta_\varepsilon - \theta)^2|Y < \varepsilon]}{\theta^3} - \cdots.$$

经整理, 可得结果. □

8.3.4 例子——均匀随机变量

由例 7.2.3 可知: 对一个在 d 维单位立方体内服从均匀分布的点的相关积分是 $F_Y(y) = y^d(2-y)^d$. 因而, 对于 $k = 1, \cdots, d$,

$$\Phi^{(k)}(y) = (-1)^k \frac{\Gamma(d+1)}{\Gamma(d-k+1)}(2-y)^{d-k}.$$

由上可知

$$r(\varepsilon) = \sum_{k=1}^{d} \frac{\Gamma(d+1)}{\Gamma(d+k+1)} \frac{\Gamma(d+1)}{\Gamma(d-k+1)} \left(\frac{\varepsilon}{2-\varepsilon}\right)^k.$$

Pisarenko 和 Pisarenko(1995) 得到了当 $q = 2$ 时, Takens 估计的方差的表达式; 同时给出了其相关性质.

8.4 Hill 估计

该估计是以 Hill(1975) 的工作为基础. 设 Y 表示 q 阶内点距离. 考虑 n 个独立内点距离的一个抽样, 设 $Y_{(m)}$ 是第 m 个顺序统计量, $1 \leqslant m \leqslant n$, i.e., 在条件 $Y_{(m)} \leqslant \varepsilon$ 下

$$Y_{(1)} \leqslant Y_{(2)} \leqslant \cdots \leqslant Y_{(m)} \leqslant \cdots \leqslant Y_{(n)}.$$

8.4.1 顺序统计量的分布

第 m 个顺序统计量的分布函数为

$$\begin{aligned} F_{Y_{(m)}}(y) &= \Pr\{Y_{(m)} \leqslant y\} \\ &= \Pr\{\text{至少存在一个 } Y \text{ 的 } m, \text{ 使得} Y_{(m)} \leqslant y\} \\ &= \sum_{r=m}^{n} \binom{n}{r} F_Y^r(y)[1 - F_Y(y)]^{n-r} \\ &= \frac{\Gamma(n+1)}{\Gamma(m)\Gamma(n-m+1)} \int_0^{F_Y(y)} t^{m-1}(1-t)^{n-m} dt \\ &= n\binom{n-1}{m-1} \int_0^{y} F_Y^{m-1}(r)[1 - F_Y(r)]^{n-m} f_Y(r) dr. \end{aligned}$$

由上得 $Y_{(m)}$ 的密度函数为

$$f_{Y_{(m)}}(y) = n\binom{n-1}{m-1} F_Y^{m-1}(y)[1 - F_Y(y)]^{n-m} f_Y(y).$$

由 David(1970, 等式 (2.2.2)), 得前 n 个顺序统计量的联合密度为

$$f_{Y_{(1)},Y_{(2)},\cdots,Y_{(m)}}(y_1, y_2, \cdots, y_m) = \frac{n!}{(n-m)!} f_Y(y_1) \cdots f_Y(y_m)[1 - F_Y(y_m)]^{n-m};$$

因此条件密度为

$$\begin{aligned} f_{Y_{(1)},Y_{(2)},\cdots,Y_{(m-1)}|Y_{(m)}}(y_1, y_2, \cdots, y_{m-1}|y_m) &= \frac{f_{Y_{(1)},Y_{(2)},\cdots,Y_{(m)}}(y_1, y_2, \cdots, y_m)}{f_{Y_{(m)}}(y_m)} \\ &= \frac{(m-1)!}{F_Y^{m-1}(y_m)} f_Y(y_1) \cdots f_Y(y_{m-1}). \end{aligned}$$

8.4.2 极大似然估计

假设相关函数为 $F_Y(y) = ay^\theta$, $y < \varepsilon$. 考虑一个样本为 n 的顺序随机变量 $Y_{(1)} \leqslant Y_{(2)} \leqslant \cdots \leqslant Y_{(m)} \leqslant \cdots \leqslant Y_{(n)}$. 定义 Θ_m 为 θ 的估计即

$$\Theta_m = \frac{m-1}{-\sum_{i=1}^{m-1} \log\left(\frac{Y_{(i)}}{Y_{(m)}}\right)}. \tag{8.4}$$

证明 考虑 n 个顺序统计量的一个抽样

$$y_{(1)} \leqslant y_{(2)} \leqslant \cdots \leqslant y_{(m)} \leqslant \cdots \leqslant y_{(n)}.$$

由于 $f_Y(y) = a\theta y^{\theta-1}$, 其条件对数似然方程为

$$L(\theta; y_1, y_2, \cdots, y_m) = \log(m-1)! - (m-1)\log(ay_{(m)}^{\theta}) + \sum_{i=1}^{m-1}\log(a\theta y_{(i)}^{\theta-1})$$

$$= \log(m-1)! - \theta(m-1)\log(y_{(m)}) + (m-1)\log\theta + (\theta-1)\sum_{i=1}^{m-1}\log y_{(i)},$$

其导数为

$$\frac{dL(\theta; y_{(1)}, \cdots, y_{(m)})}{d\theta} = \frac{m-1}{\theta} + \sum_{i=1}^{m}\log\left(\frac{y_{(i)}}{y_{(m)}}\right).$$

令上式为零, 得出估值; 然后利用常规的解法进行求解. □

8.4.3 命题

已知 $F_Y(y) = ay^{\theta}$, a 是一个正常数, 那么

$$E\left[\frac{1}{\Theta_m}\middle| Y_{(m)}\right] = \frac{1}{\theta},$$

$$E[\Theta_m | Y_{(m)}] = \frac{m-1}{m-2}\theta,$$

及

$$\text{Var}\,(\Theta_m | Y_{(m)}) = \left(\frac{m-1}{m-2}\right)^2 \frac{\theta^2}{m-3}.$$

证明.

$$E\left[\frac{1}{\Theta_m}\middle| Y_{(m)} = y_m\right]$$

$$= \int_0^{y_m}\cdots\int_0^{y_2}\frac{-1}{m-1}\sum_{i=1}^{m-1}\log\left(\frac{y_i}{y_m}\right)f_{Y_{(1)}\cdots Y_{(m-1)}|Y_{(m)}}(y_1, \cdots, y_{m-1}|y_m)dy_1\cdots dy_{m-1}$$

$$= -\int_0^{y_m}\cdots\int_0^{y_2}\frac{(m-1)!}{m-1}\sum_{i=1}^{m-1}\log\left(\frac{y_i}{y_m}\right)\left(\frac{\theta}{y_m}\right)^{m-1}\left(\frac{y_1}{y_m}\right)^{\theta-1}\left(\frac{y_{m-1}}{y_m}\right)^{\theta-1}dy_1\cdots dy_{m-1}.$$

用 $w_i = -\log(y_i/y_m)$ 进行代替, 得

$$
\begin{aligned}
&E\left[\frac{1}{\Theta_m}\middle| Y_{(m)}=y_m\right]\\
=&\frac{(m-1)!}{m-1}\int_0^\infty\cdots\int_{w_2}^\infty\left(\sum_{i=1}^{m-1}w_i\right)\theta^{m-1}\exp\left(-\theta\sum_{i=1}^{m-1}\omega_i\right)dw_1\cdots dw_{m-1}\\
=&\frac{(m-1)!}{m-1}\int_0^\infty\cdots\int_{w_2}^\infty\left(\sum_{i=1}^{m-1}i(w_i-w_{i+1})\right)\theta^{m-1}\exp\left(-\theta\sum_{i=1}^{m-1}i(w_i-w_{i+1})\right)dw_1\cdots dw_{m-1}\\
=&\frac{1}{m-1}\int_0^\infty\cdots\int_0^\infty\left(\sum_{i=1}^{m-1}z_i\right)\theta^{m-1}\exp\left(-\theta\sum_{i=1}^{m-1}z_i\right)dz_1\cdots dz_{m-1}\\
=&\frac{1}{\theta}.
\end{aligned}
$$

在第二步到最后一步,

$$z_i=i(w_i-w_{i+1}),$$

$i=1,\cdots,m-1$. 注意

$$\sum_{i=1}^{m-1}z_i=\sum_{i=1}^{m-1}w_i.$$

由定义可知 $w_m=0$. 因此

$$
\begin{aligned}
w_{m-1}&=\frac{z_{m-1}}{m-1},\\
w_{m-2}&=\frac{z_{m-2}}{m-2}+\frac{z_{m-1}}{m-1},\\
&\vdots\\
w_1&=\frac{z_1}{1}+\frac{z_2}{2}+\cdots\frac{z_{m-1}}{m-1}.
\end{aligned}
$$

故其雅可比行列式为

$$J=\left|\left(\frac{\partial w_i}{\partial z_j}\right)_{ij}\right|=\frac{1}{(m-1)!}.$$

用以上相同的替代, 得

$$
\begin{aligned}
&E[\Theta_m|Y_{(m)}]\\
=&E\left[\frac{m-1}{-\sum_{i=1}^{m-1}\log\left(\frac{Y_{(i)}}{Y_{(m)}}\right)}\middle| Y_{(m)}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^\infty \cdots \int_0^\infty \frac{m-1}{z_1+\cdots+z_{m-1}} \theta^{m-1} \exp\left(-\theta \sum_{i=1}^{m-1} z_i\right) dz_1 \cdots dz_{m-1} \\
&= E\left[\frac{m-1}{Z_1+\cdots+Z_{m-1}} | Z_i \text{ 独立同分布于参数为 } \theta \text{ 的指数分布}\right] \\
&= (m-1) \int_0^\infty \frac{1}{w} \frac{\theta(\theta w)^{m-2}}{\Gamma(m-1)} \exp(-\theta w) dw \\
&= \frac{m-1}{m-2}\theta.
\end{aligned}
$$

将相同的方法应用于 Gamma 分布, 得

$$
\mathrm{Var}[\Theta_m | Y_{(m)}] = \left(\frac{m-1}{m-2}\right)^2 \frac{\theta^2}{m-3}. \qquad \square
$$

现在应用类似的变换推导一般情况下的偏差, 此时 $F_Y(y) = \Phi(y) y^\theta$.

8.4.4 命题

已知 $F_Y(y) = \Phi(y) y^\theta$, 那么

$$
E\left[\frac{1}{\Theta_m} | Y_{(m)} = y\right] = \frac{1}{\theta}(1+\gamma_m(y)). \tag{8.5}
$$

其中,

$$
\gamma_m = E\left[\frac{1}{m-1} \sum_{i=1}^{m-1} \log\left(\frac{\Phi(Y_{(i)})}{\Phi(Y_{(m)})}\right) \middle| Y_{(m)} = y\right].
$$

进而, 若 $E[(\Theta_m - \theta)^k | Y_{(m)}] < \infty$ 对所有的 $k \geqslant 1$ 都成立, 那么

$$
E[\Theta_m | Y_{(m)} = y] = \theta(1-\gamma_m(y)) + \frac{E[(\Theta_m - \theta)^2 | Y_{(m)} = y]}{\theta} - \cdots. \tag{8.6}
$$

证明 若 $Y_{(i)}$ 是不同的第 i 阶顺序统计量, 此处 $Y_{(1)} \leqslant Y_{(2)} \leqslant \cdots \leqslant Y_{(m)}$, 那么

$$
\frac{F_Y(Y_{(1)})}{F_Y(Y_{(m)})} \leqslant \frac{F_Y(Y_{(2)})}{F_Y(Y_{(m)})} \leqslant \cdots \leqslant \frac{F_Y(Y_{(m-1)})}{F_Y(Y_{(m)})}
$$

是一个顺序统计量集合在 $[0,1]$ 上的均匀分布, 进行对数变换

$$
W_{(i)} = -\log\left(\frac{F_Y(Y_{(i)})}{F_Y(Y_{(m)})}\right),
$$

其中 $i = 1, \cdots, m$, 则顺序统计量满足

$$
W_{(1)} \geqslant W_{(2)} \geqslant \cdots \geqslant W_{(m)} = 0,
$$

其服从指数分布且均值为 1. 令 $Z_i = i(W_{(i)} - W_{(i+1)})$. Z_i 是独立同分布的指数分布随机变量构成的集合, 且其均值为 1. 进而

$$\sum_{i=1}^{m-1} Z_i = \sum_{i=1}^{m-1} W_{(i)} = -\sum_{i=1}^{m-1} \log\left(\frac{F_Y(Y_{(i)})}{F_Y(Y_{(m)})}\right),$$

因此

$$E\left[\frac{-1}{m-1}\sum_{i=0}^{m-1} \log\left(\frac{F_Y(Y_{(i)})}{F_Y(Y_{(m)})}\right)\middle| Y_{(m)}\right] = 1.$$

已知 $F_Y(y) = \Phi(y)y^\theta$, 那么

$$E\left[\frac{-\theta}{m-1}\sum_{i=0}^{m-1} \log\left(\frac{Y_{(i)}}{Y_{(m)}}\right)\middle| Y_{(m)}\right] + E\left[\frac{-1}{m-1}\sum_{i=0}^{m-1} \log\left(\frac{\Phi(Y_{(i)})}{\Phi(Y_{(m)})}\right)\middle| Y_{(m)}\right] = 1.$$

经整理, 可得方程 (8.5). 现在考虑 $1/\Theta_m$ 关于 $\dfrac{1}{\theta}$ 的泰勒展式, 取期望得

$$E\left[\frac{1}{\Theta_m}\middle| Y_{(m)}\right] = \frac{1}{\theta} - \frac{E[(\Theta_m - \theta)|Y_{(m)}]}{\theta^2} + \frac{E[(\Theta_m - \theta)^2|Y_{(m)}]}{\theta^3} - \cdots.$$

经整理, 代入方程 (8.5) 可以得到方程 (8.6). 特别需注意偏差与方差之间的关系.

8.4.5　引理

已知对任意的 $k > 1$, $E[(\Theta_m - \theta)^k|Y_{(m)}]$ 存在, 那么

$$E[(\Theta_m - \theta)^2|Y_{(m)}] = \theta^4 E\left[\left(\frac{1}{\Theta_m} - \frac{1}{\theta}\right)^2\middle| Y_{(m)}\right] + \frac{1}{\theta}E[(\Theta_m - \theta)^3|Y_{(m)}] - \cdots.$$

证明　先得出 $\dfrac{1}{\Theta_m}$ 关于 $\dfrac{1}{\theta}$ 的泰勒展式, 整理后两边平方可得

$$\left(\frac{1}{\Theta_m} - \frac{1}{\theta}\right)^2 = \left(\frac{\Theta_m - \theta}{\theta^2}\right)^2 - \frac{(\Theta_m - \theta)^3}{\theta^5} + \cdots.$$

8.4.6　$\Phi(y)$ 的单调性

如果 $\Phi(y)$ 是单调递减函数, 那么当 $Y_{(i)} \leqslant Y_{(m)}$ 时, 有 $\Phi(Y_{(i)}) \geqslant \Phi(Y_{(m)})$. 因此 $\gamma_m > 0$. 假设等式 (8.6) 中的高阶项足够小, 那么 Θ_m 低估 θ. 这与例 7.2.2 和例 7.2.3 一致, 其中的 $\Phi(y)$ 也是单调递减的, 这是由边界效应所致.

8.4.7　定理 (Mason, 1982)

若 $Y_1, Y_2, \cdots$ 是独立同分布的随机变量, 其共同的分布为 $F_Y(y)$. 假设 $m = m(n) \to \infty$, $m = o(n)$. 那么 Θ_m 是 θ 的一个弱估计当且仅当 $F_Y(y)$ 有一个指数为 θ 正则变化的下尾 (定义 7.4.2). Mikosch 和 Wang 进一步证明了当 $m = n^\alpha$ 时, 对某些 $\alpha \in (0,1)$, Θ_m 是强一致渐近于正态分布.

8.4.8 推论

在定理 8.4.7 的条件下, Θ_m 是 θ 的一个弱一致估计当且仅当 $\Phi(y)$ 缓慢的趋向于 0.

证明 由定理 8.4.7 和命题 7.4.5 可得.

根据推论 8.4.8, 如果 $F_Y(y)$ 是缺项的, 那么 Θ_m 不是 θ 的一致估计. 这使得相比 Grassberger-Procaccia 估计来说, Θ_m 估计严重有偏. 同样地, Takens 估计也严重有偏. Smith(1992a) 在 β 二项分布的基础上对于 Takens 估计进行了修正, 这种修正使得估计偏差变小了.

得出现在和以前章节的结论的前提是假设样本的点间距离是独立的. 在一般系统中使用这种方法仍然是出于经验. 一些作者提议使用自举过程, 即随机选取相对较少的已观测到的点间距离, 作为加强被选点间距离的独立性的一种方法.

8.5 自举估计过程

Grassberger-Procaccia(1983a) 提出的 D_2 估计方法是关于如何决定所有点间距离经验概率分布幂律指数的. Takens 和 Hill 估计是基于独立点间距离的随机样本, 并且都涉及到了充分小距离的对数和. 即使只有少量的小的点间距离也能够通过 Grassberger-Procaccia 方法进行估计, 但 Hill 估计 (或者 Takens 估计 Θ_ε)Θ_m 对于小的 m(或者 ε) 来说, 估计很不稳定.

Mikosch 和 Wang 提出利用自举抽样过程来减小样本的方差. 这样的自举过程提供了维数估计标准差的粗略估计. 在 1993 年, Takens 还提出了 Monte Carlo 模拟. 自举技术被 Efron 和 Tibshirani(1986, 1993) 广泛地讨论.

8.5.1 例子——Cantor 测度

考虑例 7.3 中关于 Cantor 测度的讨论, 特别是, 当 $p = 0.5$ 时的情况. 一个 $N = 100,000$ 的在 Cantor 测度下具有概率分布的随机样本, 能通过生成 0 和 2 的二元序列来模拟, 并且转化为二进制下 0 和 1 之间的一个小数.

D_2 的五种 Hill 估计都已经用等式 (8.4) 的估计值进行了计算. 每个估计都是基于 $n = 100,000$ 点间距离的随机样本, 并且 Hill 估计对于 m 的任意取值都已经进行了计算. m 的取值决定了点间距离 $y_{(m)}$ 的最大顺序统计量, m 的值是点间距离的一个替代, 所以 m 关于 $\hat{\theta}_m$ 的图相似于由 $\hat{\theta}_m$ 确定的某一点间距的图. 可选取 m 取值的范围使得 $y_{(m)}$ 生成具有幂律特征的点间距离. 估计越多的 m 的中间值将会对曲线具有更大的决定性. 替代地, 可以画 $y_{(m)}$ 关于 $\hat{\theta}_m$ 或 $\log y_{(m)}$ 关于 $\hat{\theta}_m$ 的图. 在现有例题中, 在对数尺度下, 循环行为具有周期性, 因而如果 $F_Y(y)$ 是缺项的, 那么后者更适合.

结果见图 8.1. 值得注意的是对 $y_{(m)}$ 的较小值, 图形变化很大. 并且由于估计值的性质, 当 m 增加时, 在单个游程 (路径) 中, 存在极大的自相关性. 由图知单个游程 (路径) 的初始值要么极高要么极低, 但经过多次循环后, 它必收敛到 Θ_m 的期望值. 一个改进较小 m 值估计的方法是增加 n 的样本数对. 如果 n 的值确定了, 那么 $y_{(m)} \ll 3^{-11}$ 的值也确定了. 因而, 一旦 $y_{(m)} > 3^{-11}$, 估计会更稳定. 上述计算是困难的, 为了抽样更多的缺项循环, 数对 (n, N) 的数量至少成倍增加, 这样可在计算机上很快达到极限.

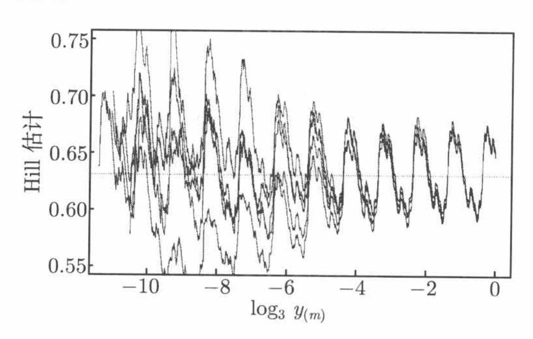


图 8.1 $p_0 = 0.5$ 时, Cantor 测度的 D_2 变化

参数 $p_0 = 0.5$ 时, Cantor 测度的 D_2 估计, 针对了五个独立的序列 ($n = 100,000$, $N = 100,000$). 虚线表示 D_2

另一种方法是重复估计过程多次. 正如图 8.1 所示, 对 m 的每一取值, 计算维数估计的平均值. 这正是自举估计方法, 这种方法独特的优势是每个 m 的维数估值的标准差可以粗略地进行估计. 我们估计 $D_q(q = 2, 3, \cdots)$ 的自举过程不同于 Mikosch 和 Wang(1995) 算法. 下面例题将予以说明.

8.5.2 例子—— 指数分布

Takens 和 Hill 估计是基于假设内点间间距服从幂律分布. 这一假设并不完全正确, 特别是有缺项的时候. 然而, 即使是缺项情形, 幂律分布也是很好的首选近似; 例如图 7.4. 如果分布正好是幂律分布, 那么曲线将变为一条直线, 该直线的斜率恰为幂指数. 震荡行为是缺项的具体显现. 服从幂律分布的随机变量其对数服从指数分布. 在本例中简单的考虑这种特殊情况.

假设 Z 是一个参数为 λ 的指数随机变量. 那么 $E[Z] = \lambda^{-1}$, $\mathrm{Var}(Z) = \lambda^{-2}$. 考虑 k 个均值的自举样本估计, 样本大小为 m. 我们特别感兴趣的是当 m 很小 (< 10), 但 $k \approx 100$ 时的情况. 记独立同分布的均值的估计值为 $\bar{Z}_1, \cdots, \bar{Z}_k$. 现在考虑 λ 的两个估计值,

$$\Lambda_1 = \frac{1}{k}\left(\frac{1}{\bar{Z}_1} + \cdots + \frac{1}{\bar{Z}_k}\right),$$

$$\Lambda_2 = \frac{k}{\bar{Z}_1 + \cdots + \bar{Z}_k} = \frac{1}{\bar{Z}^*},$$

这里 $\bar{Z}^*$ 是 $m \times k$ 个独立同分布随机变量的样本均值. 如果所有的样本被看做一个整体, $\bar{Z}^*$ 就是它的最大似然估计值. 因为

$$E\left[\frac{1}{\bar{Z}_1}\right] = \frac{m}{m-1}\lambda, \quad \mathrm{Var}\left[\frac{1}{\bar{Z}_1}\right] = \left(\frac{m}{m-1}\right)^2 \frac{\lambda^2}{m-2},$$

所以

$$E[\Lambda_1] = \frac{m}{m-1}\lambda, \quad E[\Lambda_2] = \frac{mk}{mk-1}\lambda.$$

类似地,

$$\mathrm{Var}(\Lambda_1) = \frac{1}{k}\left(\frac{m}{m-1}\right)^2 \frac{\lambda^2}{m-2}, \quad \mathrm{Var}(\Lambda_2) = \frac{1}{k}\left(\frac{mk}{mk-1}\right)^2 \frac{\lambda^2}{mk-2}.$$

因此, 可以看出, Λ_2 不仅偏差很小, 而且方差也很小. 特别地, 当 $k = 100, m = 5$ 时,

$$E[\Lambda_1] = \frac{5\lambda}{4}, \quad E[\Lambda_2] = \frac{500\lambda}{499}, \quad \mathrm{Var}(\Lambda_1) \approx 2.5 \times \mathrm{Var}(\Lambda_2).$$

8.5.3 自举化的 Hill 估计

现在建立一个抽样框架, 这将在第 9 章得到充分应用. 设 $x_1, \cdots, x_N$ 是 N 个 d 维向量观察值序列. 并且设 $A_0 = \{x_1, \cdots, x_N\}$, A_ε 是除了靠近边界观察值以外点的集合, 因此 $A_\varepsilon \subset A_0$. 不靠近边界观察值的个数由 ε 的大小决定, $0 \leqslant \varepsilon \leqslant \frac{1}{2}$.

每个自举样本都用到下面的过程. 设 i 代表当前自举样本标号. 在每个自举样本间有 n 个样本内点间距; 即 y_s, 此处 $s = 1, \cdots, n$.

1. 对于每个 s, 样本 $x_{s_q} \in A_\varepsilon$ 及 $x_{s_1}, x_{s_2}, \cdots, x_{s_{q-1}} \in A_0$. 设

$$y_s = \max\{\|x_{s_1} - x_{s_q}\|, \cdots, \|x_{s_{q-1}} - x_{s_q}\|\}.$$

2. 整理得到顺序统计量: $y_{(1)} < \cdots < y_{(m)} < \cdots < y_{(n)}$.
3. 对于 m 取不同的值, 计算

$$z_{i,m} = \frac{-1}{m-1} \sum_{j=1}^{m-1} \log\left(\frac{y(j)}{y(m)}\right). \tag{8.7}$$

重复上面的步骤 k 次, 其中 k 是自举样本的个数. 于是自举化的 Hill 估计为

$$\hat{\theta}_m = \frac{k}{\sum_{i=1}^{k} z_{i,m}}. \tag{8.8}$$

设 $\hat{y}_{(m)}$ 是所有 k 个自举样本中第 m 个顺序统计量的均值. 取定值 m, 使得中点距离 $\hat{y}_{(m)}$ 扩张到 $F_Y(y)$ 幂律行为所属的范围. 中间点中较大的点将赋予较大的曲率. 可能的图形包括 m 关于 $\hat{\theta}_m$, $\bar{y}_{(m)}$ 关于 $\hat{\theta}_m$ 及 $\log_b \bar{y}_{(m)}$ 关于 $\hat{\theta}_m$ 的图形, 此处 b 是恰当的基.

在 Grassberger-Procaccia 方法 (8.2 节) 中, 可以通过 $\log \hat{F}_Y^{GP}(y, N)$ 的样本值画 $\log y$ 的图形; 对于给定的 $\hat{\theta}$, 计算直线的斜率. 在上面的 Hill 方法中, $\hat{\theta}$ 可以从纵轴上直接看出. 在缺项情形中, 决定在哪画出这条直线很不容易. 粗略地说, Hill 估计的图实质上是 $\log \Phi(y)$ 的一个以 $\hat{\theta}$ 为中心的缩放形式. 这些难点将在 8.6 节进一步讨论.

8.5.4　标准差

自举过程涉及两个层次的取样. 开始的样本是来自 χ, 记为 $x_1, \cdots, x_N$. $x_i \in \chi, i = 1, \cdots, N$ 按经验分布取每个点 x_i 的概率都是 $\dfrac{1}{N}$. 第二个层次是自举过程, 它涉及从经验分布中重复取样. 下面就是基于经验分布条件下的标准差的计算.

设 $z_{i,m}$ 是随机变量 $Z_{i,m}, i = 1, \cdots, k$ 的观察值, k 是自举样本个数. 假设随机变量 $Z_{1,m}, \cdots, Z_{k,m}$ 是独立同分布的, 且与 $\dfrac{1}{\Theta_m}$ 有相同分布. 把 θ 的自举估计值 (k 次迭代后) 记作 $\Theta_m^{(k)}$,

$$\Theta_m^{(k)} = \frac{k}{\displaystyle\sum_{i=1}^{k} Z_{i.m}},$$

并设 $\theta_m = E[\Theta_m^{(k)}]$. 对 $\mathrm{Var}(Z_{1,m})$ 进行估计, 它是自举方法的结果, 现在导出 $\mathrm{Var}(\Theta_m^{(k)})$ 的估计值. 用 $\Theta_m^{(k)}$ 和 θ_m 分别代替 Θ_m 和 θ, 引理 8.4.5 仍然成立, 即

$$E[(\Theta_m^{(k)} - \theta_m)^2 | Y_{(m)}] = \theta_m^4 E\left[\left(\frac{1}{\Theta_m^{(k)}} - \frac{1}{\theta_m}\right)^2 \middle| Y_{(m)}\right] + \frac{1}{\theta_m} E[(\Theta_m^{(k)} - \theta)^3 | Y_{(m)}] - \cdots .$$

等式右边第二项用泰勒公式展开, 首项涉及到 $E[(1/\Theta_m^{(k)} - 1/\theta_m)^3]$. 因此, 假设自举样本独立,

$$E[(\Theta_m^{(k)} - \theta_m)^2 | Y_{(m)}] = \theta_m^4 \mathrm{Var}\left(\frac{Z_{1,m} + \cdots + Z_{k,m}}{k}\right) + o(k^{-2}) = \frac{\theta_m^4}{k} \mathrm{Var}(Z_{1,m}) + o(k^{-2}).$$

$\mathrm{Var}(Z_{1,m})$ 的估计为

$$S_{Z_{1,m}}^2 = \frac{1}{k-1} \sum_{i=1}^{k} Z_{i,m}^2 - \frac{k}{k-1} \left[\frac{1}{k} \sum_{i=1}^{k} Z_{i,m}\right]^2 . \tag{8.9}$$

因此

$$S_{\Theta_m^{(k)}}^2 = \frac{\hat{\theta}_m^4}{k} S_{Z_{1,m}}^2 \tag{8.10}$$

是 $\mathrm{Var}(\Theta_m^{(k)})$ 的估计, 此处 $\hat{\theta}_m$ 和 $S_{Z_{1,m}}$ 分别由等式 (8.8) 和 (8.9) 给出.

8.6 讨论和例子

正如推论 8.4.8 所述, 当 $F_Y(y)$ 有缺项时, Hill 估计并不一致. 本节, 首先考虑两个例子: 当推论 8.4.8 的条件满足时的均匀分布; 当推论 8.4.8 的条件不满足时的 Cantor 测度. 我们将试着去找 Hill 估计偏差的来源; 并将缺乏一致性的 Hill 估计与 Grassberger-Procaccia 方法进行对比.

8.6.1 均匀分布的例子

在 $[0,1]$ 上服从均匀分布的样本 $(N = 10,000)$ 已被模拟, 并且运用等式 (8.8) 计算出 D_2 的 Hill 估计, 图 8.2 标绘出了各种 m 的值. 这 $k = 100$ 自举样本中的每一个样本均由 $n = 10000$ 的内点距离构成的, 横坐标为 $\log_{10} \bar{y}_{(m)}$, 是第 m 个顺序统计量的所有 $k = 100$ 自举样本的平均, 尤其是有缺项情况的时候, 维数估计常是关于内点间距离的对数画图. 用等式 (8.10) 计算, 并叠加两倍于标准误差的时间间隔.

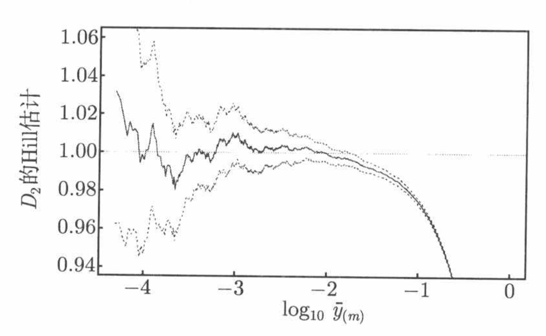


图 8.2 一致分布的 D_2 估计

一致分布的 D_2 估计（$k = 100$, $n = 100,000$, $N = 100,000$). 上、下短划线标记了由两个标准误差给出的区间宽度

由 §7.2.3, $F_Y(y) = y^d(2-y)^d$, 因此 $\Phi(y) = (2-y)^d$ 在零点处缓慢地变化, 这里 $d = 1$. 由推论 8.4.8 知, 估计是一致的, 且渐近于无偏估计. 然而, 这些估计受 §7.2 中所讨论的边界效应的影响, 这将在 §9.2 中作进一步的讨论. 为了得到一个合理精确的维数估计, 样本必须是足够小的内点间距离, 这样函数 $F_Y(y)$ 将被幂律部分所控制, 并且 $\Phi(y)$ 部分对其影响甚小.

在高维场合, 抽样小的内点间距的似然律将会下降. 见 $\mathbb{R}^d$ 上白噪声图 10.13 中 D_2 维数的估计, 其中 $d = 1, \cdots, 5$. $d = 1$ 时, 幂律行为是明显的, 但是随着 d 的增加, 边界效应增强, 并且 $\Phi(y)$ 的增大将起主导作用, 这并不是 Hill 估计特有的.

这将在 §9.2 中进行讨论.

现在考虑一个例子, 它的特征与前面讨论的例子相反. 由 Cantor 测度得到的概率分布 $F_Y(y)$ 是有缺项的. 因此, 根据推论 8.4.8, Hill 估计不是一致的. 然而, 在这个例子中, 一种幂律行为延伸到了整个单位区间 (见定义 7.4.6), 边界效应对它的影响方式不同于对前面例题的影响方式.

8.6.2　Cantor 测度的例子

一个 $N = 100,000$ 的随机样本在例 8.5.1 中已被模拟, 概率分布由 Cantor 测度给出. 图 8.3 和图 8.4 描绘了 $p_0 = 0.5, 0.2$ 时, $\hat{\theta}_m$ 的估计 (公式 (8.8)). 用 m 和 $\log_3 \bar{y}_{(m)}$ 描绘出估计, 其中 $\bar{y}_{(m)}$ 是第 m 个顺序统计量的所有 $k = 100$ 个自举样本的平均, 并对 $n = 100,000$ 内点距离进行抽样. 图 8.3 和图 8.4 中虚线代表着 D_2 的真实值, 这可以由定理 5.5.4 和等式 2.3 及 3.8 算出来. 根据推论 8.4.8, 在这个例子中 Hill 估计并不是一致的, 这在图 8.3 和图 8.4 中可以明显看出来.

注意到当 $\bar{y}_{(m)} \to 0$(底部图) 时, 由等式 (8.10) 给出的标准误差增加. 该标准误差的图像大致是线性的. 因此当 $\bar{y}_{(m)} \to 0$ 时, 幂增加, 它和图 8.4 一样也包含缺项, 这在图 8.4 中更加明显. 当内点距离是一个固定的数时, 如果当 $p_0 = 0.2$, 即测度不是均匀分布时, 更小的值将被选择. 因此, 在此情形下我们能观察更多的缺项周期.

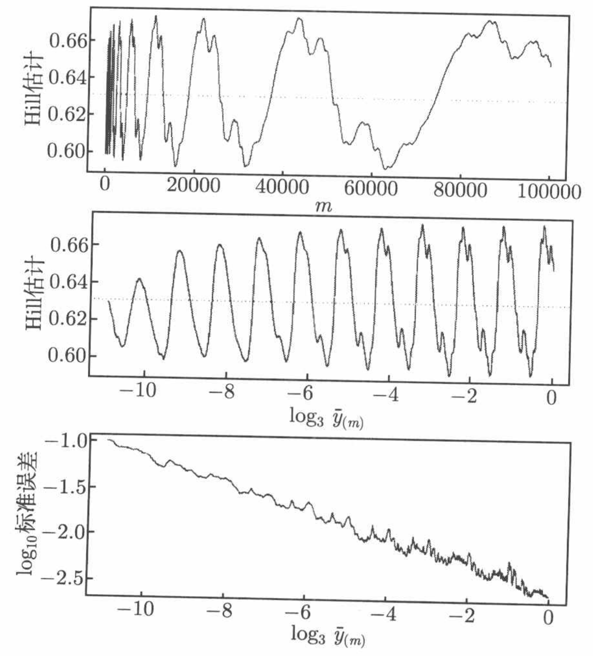


图 8.3　$p_0 = 0.5$ 时, Cantor 测度的 D_2 估计

参数 $p_0 = 0.5$ 时, Cantor 测度的 D_2 估计 ($k = 100$, $n = 100,000$, $N = 100,000$). 虚线表示 D_2

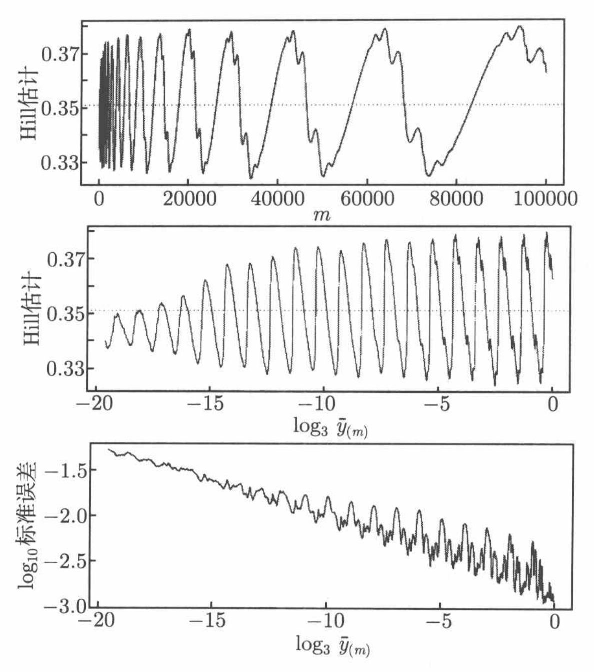


图 8.4 $p_0 = 0.2$ 时, Cantor 测度的 D_2 估计

参数 $p_0 = 0.2$ 时, Cantor 测度的 D_2 估计 ($k = 100$, $n = 100,000$, $N = 100,000$). 虚线表示 D_2

通过这些图, 我们提出一个问题: 代表 D_q 的估计线应画在何处? 第一个近似值可能把代表 D_q 的线放在中间, 即放在缺项周期的顶峰和低谷之间. 然而, 不久将知道 D_5 是个特殊的情况, 它很明显地不位于缺项周期的中间. 事实上, 在所有的这些情形中, 用"中间"作为 D_q 的估计将导致正的偏差.

8.6.3 在何处画线

为了确定在何处画线, 尤其是存在缺项时, 返回到等式 (8.6), 它给出了 Hill 估计期望值, 即

$$E[\Theta_m|Y_{(m)} = y] = \theta[1 - \gamma_m(y)] + \frac{E[(\Theta_m - \theta)^2|Y_{(m)} = y]}{\theta} - \cdots, \tag{8.11}$$

$$\gamma_m(y) = E\left[\frac{1}{m-1}\sum_{i=1}^{m-1}\log\left(\frac{\Phi(Y_{(i)})}{\Phi(Y_{(m)})}\right)\Bigg| Y_{(m)} = y\right].$$

现在假设当 m 充分大时,

$$\gamma_m(y) \approx E[\log\Phi(Y)|Y < y] - \log\Phi(y),$$

等式 (8.11) 中的高阶项忽略不计. 如果方程 (8.11) 中的高阶项可以略掉, 那么

$$E[\Theta_m|Y_m = y] \approx \theta[1-\gamma_m(y)] \approx \theta + \theta\log\Phi(y) - \theta E[\log\Phi(Y)|Y<y]. \tag{8.12}$$

我们相信: 如果 $F_Y(y)$ 是缺项的, 那么对任意的 y 值. 当 $y \to 0$ 时, $E[\log\Phi(Y)|Y < y]$ 将包含无限多个循环. 因此, $E[\log\Phi(Y)|Y<y]$ 将达到一个极限常数. 不幸的是, 情况并不是这样. 期望取的是对应于 Y 的概率分布的期望. 由假设得, 这个期望是幂律的. 大的内点间距离增加的可能性使期望与 $\log\Phi(y)$ 有相似的周期行为, 尽管振幅很小. 因此, 我们期望维数估计本质上是 $\log\Phi(y)$ 的一个估计, $\log\Phi(y)$ 以 θ 为中心, 但是中心由于周期性会轻微移动. 从图 8.3 可以看出这一点. 它至少表面上看起来是图 7.5 中 $\log\Phi(y)$ 的尺度变化和位移.

在何处画线是通过考虑方程 (8.12) 的条件期望得到的. 特别地,

$$E[E[\Theta_m|Y_{(m)}=y]|Y<y] \approx \theta + \theta E[\log\Phi(y)|Y<y] - \theta E[E[\log\Phi(Y)|Y<Y']|Y'<y]. \tag{8.13}$$

在图 8.3 和图 8.4 的样本关系上, 它等价于在维数估计时计算部分均值, 进而有效地修匀缺项循环. 由下面的例子知, 诱使方程 (8.12) 中的偏倚的两项之间的差异将在方程 (8.13) 中被大大地缩小. 重复取条件期望 (即修匀), 假设维数估计充分稳定, 偏倚项的影响将被进一步降低.

8.6.4 Cantor 例子的继续

进一步, Cantor 测度的 $D_2, \cdots, D_5$ 估计可在图 8.5 和图 8.6 中找到. 在这两种情况下, 与图 8.3 和图 8.4 中一样, 产生了 $N = 100,000$ 的样本序列. 在 $k = 100$ 的自举样本中, 每一个都包含 $n = 100,000$ 个 q 阶内点距离. 振幅最大的线是由方程 (8.8) 给出的未修匀的自举估计. 中等振幅的线表示部分均值 (即一次修匀), 这由方程 (8.3) 给出. 最小振幅的线表示经过两次修匀后的 Hill 估计. 进一步修匀就进一步减少震荡行为. 值得注意的是, 修匀估计并不稳定, 除非 $\bar{y}_{(m)}$ 变得相当大. 这种不平稳性是由原始的 Hill 估计所引起的. 在修匀的每个水平上, 这种不稳定性似乎向右延伸到 $\bar{y}_{(m)}$ 的更高值. 因此, 在可实施的修匀个数上有某种实际限制.

在这两幅图中, 比较大的内点距离, 修匀线振动非常接近于 Rényi 维数的真实值. 然而, 修匀估计中仍有一个正的偏倚.

8.6.5 偏倚减少

在缺项循环明确定义的情况下, Rényi 维数的初始估计可由最后一个缺项循环的修匀估计均值计算得到. 从图 8.5 和图 8.6 中可以看到, 这将产生一个带正偏倚的估计, 这个数值被作为不相关估计制成表格 8.1.

方程 (8.11) 中, 首先要排除的项涉及 $E[(\Theta_m-\theta)^2|Y(m)]$. 注意到

$$E[(\Theta_m-\theta)^2] = E[(\Theta_m - E[\Theta_m])^2] + (E[\Theta_m]-\theta)^2,$$

其中每一个期望都是在 $Y_{(m)}$ 上的条件期望. $\mathrm{Var}(\Theta_m)$ 的一个估计是 $S^2_{\Theta(m)}$, 自举样本方差由方程 (8.10) 给出. 然而, 这一项比起 $(E[\Theta_m]-\theta)^2$ 来相对较小, 所以这一点可以忽略不计. 这意味着比方程 (8.12) 中一个更好的逼近是

$$\begin{aligned}&E[\Theta_m|Y_{(m)}=y]-\frac{(E[\Theta_m]-\theta)^2}{\theta}\\ \approx&\theta+\theta\log\Phi(y)-\theta E[\log\Phi(Y)|Y<y].\end{aligned}$$

因此, D_q 的一个"正确的"估计可由原始 Hill 估计 (方程 (8.8)) 减去方差项得到, 就像上面方程的左边一样; 但如果用不相关估计代替 θ, 那么同样进行了两步修匀.

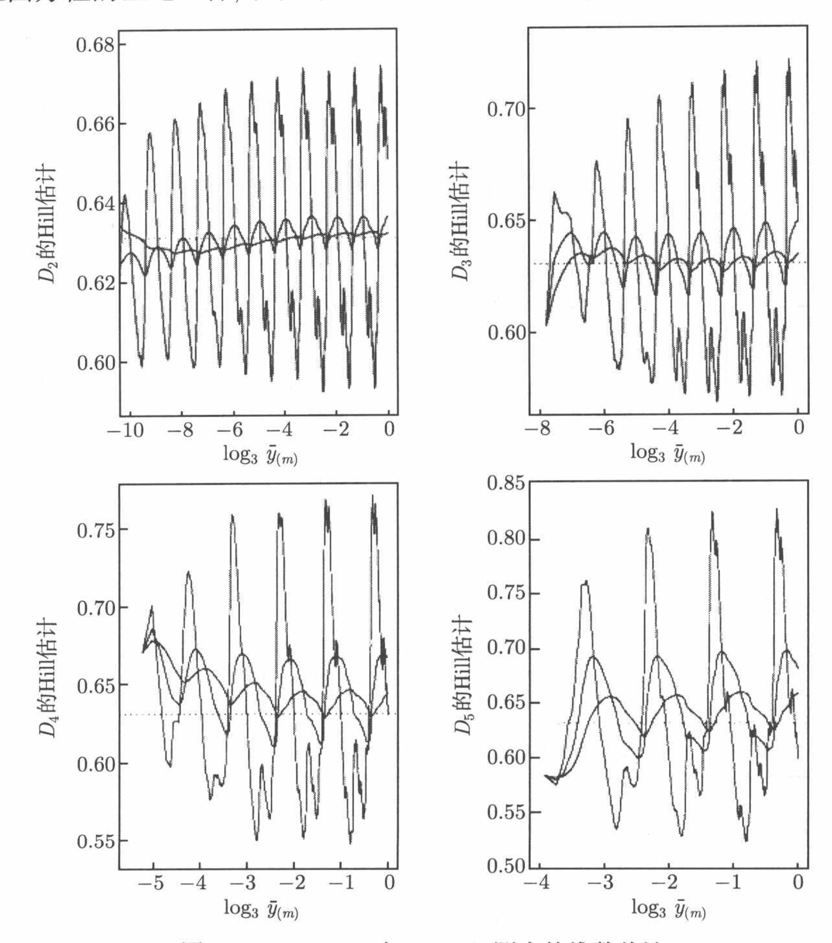


图 8.5 $p_0=0.5$ 时, Cantor 测度的维数估计

参数 $p_0=0.5$ 时, Cantor 测度的 $D_2,\cdots,D_5$ 估计 ($k=100, n=100,000, N=100,000$). 振幅最大的线是由方程 (8.8) 给出的未修匀估计, 中等振幅的线表示经过一次修匀, 振幅最小的线表示经过两次修匀. 水平虚线表示已知的 Rényi 维数 D_q

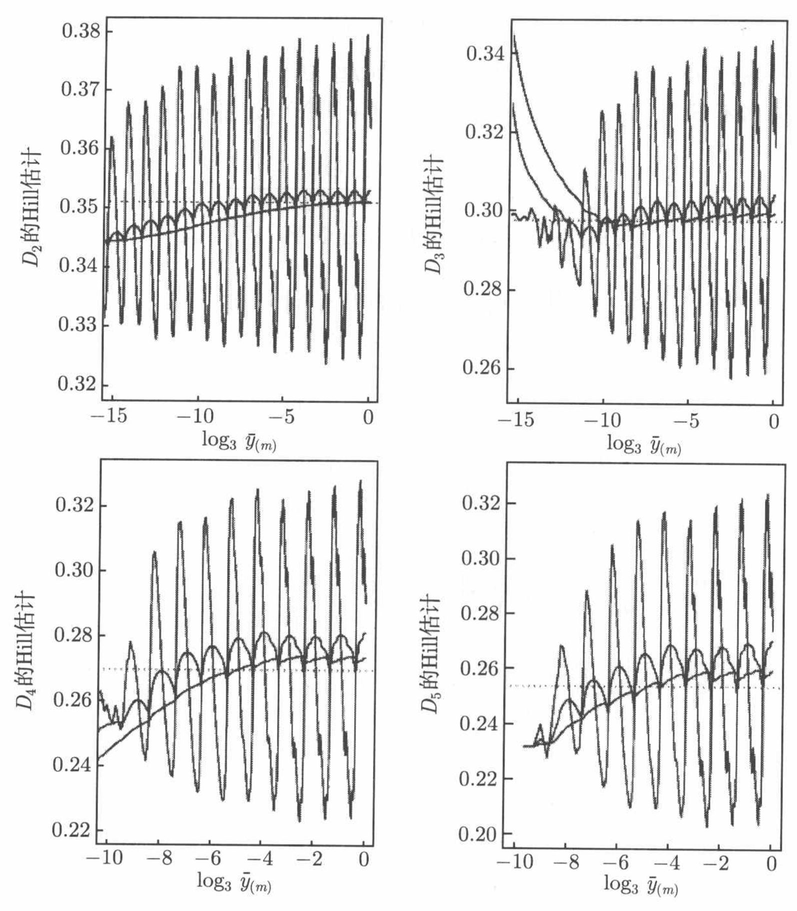


图 8.6　$p_0 = 0.2$ 时, Cantor 测度的维数估计

参数 $p_0 = 0.2$ 时, Cantor 测度的 $D_2, \cdots, D_5$ 估计 ($k = 100, n = 100,000, N = 100,000$). 振幅最大的线是由方程 (8.8) 给出的未修匀估计, 中等振幅的线表示经过一次修匀, 振幅最小的线表示经过两次修匀. 水平虚线表示已知的 Rényi 维数 D_q

表 8.1　$p_0 = 0.5$ 的 Cantor 测度, $p_0 = 0.2$ 的 Cantor 测度

$p_0 = 0.5$				$p_0 = 0.2$			
q	实际值	无关估计	相关估计	q	实际值	无关估计	相关估计
2	0.6309	0.6320	0.6310	2	0.3510	0.3512	0.3503
3	0.6309	0.6330	0.6294	3	0.2976	0.2992	0.2696
4	0.6309	0.6365	0.6295	4	0.2696	0.2730	0.2689
5	0.6309	0.6406	0.6290	5	0.2537	0.2584	0.2527
6	0.6309	0.6515	0.6349	6	0.2437	0.2504	0.2428
7	0.6309	0.6670	0.6451	7	0.2370	0.2446	0.2353
8	0.6309	0.6886	0.6618	8	0.2321	0.2413	0.2301

注: Rényi 维数 D_q 的 Cantor 测度估计. 在相关和不相关估计中, 如在图 8.5 和图 8.6 中所示.

我们应像在方程 (8.13) 中那样进行修匀. 在表 8.1 中, 校正估计是对最近的缺项循环中的被修匀和校正的 Hill 估计进行平均.

如果非幂律部分 $\Phi(y)$ 不仅包含缺项部分, 而且具有边界效应, 那么这种情况更复杂, 相关的例子将在 §10.2 中给出.

8.6.6 讨论

前面已经讨论了估计 Rényi 维数的两个方法. 一种基于最大似然法 (Takens 和 Hill), 另一种基于 log-log 对数图 (Grassberger-Procaccia) 斜率法. 这两种方法有问题, 也有相对优势. 下面简要地总结它们各自的优点和缺点.

这两种方法都可能受到边界效应的影响, 这将在第 9 章进一步讨论. 边界效应即使存在也并不总会发生. 它会导致较大的内点距离, 其大小相当于该区域的宽度. 也就是说, 幂关系对于大的相当于区域宽度的内点距离并不成立. 这两种方法也受到由较小样本量造成的更小的内点距离更大的变化的影响.

估计相关指数的 Grassberger-Procaccia 方法包括绘制 $\log y$ 关于 $\log \hat{F}_Y^{GP}(y,N)$ 图形. 通过估计斜率给出了相关指数的估计. 通常用一些最小二乘回归形式来估计斜率. 然而, 有一些需要进行任意选择. 需要选择这样的一个区间：线是充分直的, 它的上端不受边界效应的影响, 它的下端可能具有极端的变异性和缺乏数据. 另一问题是由计数的有效和累积拟合一条直线. 这样进行回归的点并不具有独立性. 并且由于不同的样本量, 有不同的方差.

Takens 和 Hill 估计是基于极大似然理论, 因而具有与这一理论相关的优美的性质. 这些估计的另一优势是它们可以被精确地定义, 因而更容易使用它们特有的性质进行分析. 然而, 当在某一内点间距范围外对所要求的 Rényi 维数进行估计, 在这些点中的每一处估计偏差和方差时, 同样会遇到 Grassberger-Procaccia 方法中所遇到的边界效应和在下端的变异性. 相对于 Grassberger-Procaccia 方法, Takens 和 Hill 估计法的优点是最终结果的图是直接的维数估计. 在 Grassberger-Procaccia 方法中, 需要计算斜率或局部导数, 这具有本质性的困难. 基于 Hill 方法的图强调缺项行为, 这如分形维数一样有趣. 它不仅告诉我们类自相似集的测度, 而且还告诉我们标度因子.

Grassberger-Procaccia 方法的优点是它是一致的, 而 Takens 和 Hill 方法并不是一致的. 假设有大量的样本, 并且 $\log y$ 关于 $\log \hat{F}_Y^{GP}(y,N)$ 的图已根据这种方法绘制. 在 Cantor 测度的场合, 图形看起来像图 7.4. 如果继续增加样本量, 抽样将更小, 内点间距将更小. 同时随着进一步的缺项循环, 图形进一步向左扩展. 当线进一步向左扩展时, 随着缺项循环的增加, 拟合直线的斜率会越来越接近真实直线的斜率. 或者, 从图 8.3 和图 8.4 也可看出, 在 Hill 方法中, 尽管我们并不清楚应该在何处画线, 随着样本量的增加, 对更小的 $\bar{y}_{(m)}$, 缺项循环进一步增加.

不过这或许不是一个问题. 这一节使用的是 Cantor 测度, 通过 Hill 方法合理精确地对维数进行了估算. 并没有将典型方法用于决定一条线“直”的部分. 从我们的数值研究发现, 如果分析的测度像 Cantor 测度那样有“gap”, 那么除了等式 (8.11) 的情况, 在等式 (8.13) 的条件下, 期望将降低 Hill 估计的振幅, 使之更接近 θ 的值. “gap”似乎保证了 $\Phi(y)$ 不包含全部的边界 (见例 10.2.1).

第 9 章　偏差的外在来源

9.1　引　　言

在第 7 章中, 已经证明了, 如果相关指数 $\theta(q)$ 存在, 则概率函数 $F_Y(y)$ 能分解成 $\Phi(y)y^{\theta(q)}$, 其中 $\Phi(y)$ 表示非幂率部分. 函数 $\Phi(y)$ 诱导的非幂率行为称为偏差的内在形式. 在第 8 章中, 探讨了估计 $D_q = \dfrac{\theta(q)}{q-1}$ 的方法, 其中 $q = 2, 3, \cdots$.

因为实验的局限或缺陷, 抽样的点间距的经验概率分布与式子 (7.6) 给出的 $F_Y(y)$ 的分布有非常不同的特性. 这里指的不是与第 8 章中使用的特定估计相关的那些偏差来源, 这里指的偏差的外在来源类似于样本调查方法中的非抽样误差, 是数据内在的一部分, 而不是基本过程的特性, 即不是内在的偏差.

在许多现实状况中, 这种数据上的缺陷能够被极小化, 但可能无法消除. 许多数据集来自于多年的观测, 在对它们进行分析之前的很长一段时间里, 就曾经被研究过了. 例如, 地震位置的记载. 地震是地球这一个大系统事件的一部分, 但大多在地球板块的边界上或靠近边界发生. 要检测某个等级以上的所有地震, 需要地震充分接近某个活跃的地震网. 通过地震网就能有效地确定边界. 在那些能够检测到地震的区域内, 地震发生位置的精度也经常变化. 然而, 地震网能连续升级, 于是边界和位置误差 (噪声) 也连续改变.

偏差的外因对维数估计的影响很难从分析上量化, 这一章主要探索三种外在的偏差：强加的边界的影响、四舍五入的影响, 以及数据中噪声的影响. 当数据模拟好的时候, 用带自举抽样程序 (§8.5) 的 Hill 估计来说明它们在维数估计上的影响.

9.2　强加的边界的影响

第 8 章中估计 Rényi 维数的方法是基于点间距. 当强加一个人工的边界时, 仅仅由边界确定的区域内的数据被收集了, 而排除了区域外的数据, 于是边界的影响就出现了.

例如当 $q = 2$ 时的点间距. 为了得到一个能与区域的宽度可比的点间距, 两个点必须从临近相对边界的地方抽取. 反之, 很有可能抽取到更小的点间距. 结果, 在没有强加边界的情形, 会缺少大的点间距.

这一节描述了通过观测比事件正在发生的区域更小的范围产生的影响, 也就是

在进行数据采样时, 强加一个边界的影响. 这些强加的边界在概念上不同于事件内在的边界, 正如第 7 章描述的那样, 但它们有同样的影响. 我们也给出了边界影响的一部分纠正.

9.2.1 边界影响的纠正

调整边界影响的一个方法就是从观测区域的一个子区域中进行子抽样. 设 $\mathcal{A}_0$ 是观测区域中的所有观测值, 即

$$\mathcal{A}_0 = \{x_j : j = 1, \cdots, N\},$$

其中 $x_j \in \mathbb{R}^d$. 令 $\mathcal{A}_\varepsilon$ 是一个更小的点集且 $\mathcal{A}_\varepsilon \subset \mathcal{A}_0$. 临近 $\mathcal{A}_0$ 的边界的一些观测从 $\mathcal{A}_\varepsilon$ 中排除掉. 下面定义 $\mathcal{A}_\varepsilon$.

令 $x_j = (x_{j1}, \cdots, x_{jd})$. 点在第 u 维空间的值域表示为

$$r_u = \max_j(x_{ju}) - \min_j(x_{ju}), \quad u = 1, \cdots, d.$$

然后定义 $\mathcal{A}_\varepsilon$ 为点的可约集, 其中 $0 \leqslant \varepsilon < \dfrac{1}{2}$,

$$\mathcal{A}_\varepsilon = \left\{ x_j : x_j \in \prod_{u=1}^{d} [\min_j(x_{ju}) + r_u\varepsilon, \max_j(x_{ju}) - r_u\varepsilon], j = 1, \cdots, N \right\}. \tag{9.1}$$

进行 §8.5 中的抽样程序. 令第 $s(s = 1, \cdots, n)$ 个 q 阶点间距为

$$y_s = \max\{\|x_{s_1} - x_{s_q}\|, \|x_{s_2} - x_{s_q}\|, \cdots, \|x_{s_{q-1}} - x_{s_q}\|\}.$$

从 $\mathcal{A}_\varepsilon$ 中抽取 x_{s_q}, 从 $\mathcal{A}_0$ 中抽取 $x_{s_1}, x_{s_2}, \cdots, x_{s_{q-1}}$, 相对于比 $\mathcal{A}_\varepsilon$ 的内区域宽度更小的点间距, 边界的影响并不是很令人注意.

9.2.2 例子——一致分布

图 9.1 表示了单位区间 [0,1] 上一致分布点的 D_2 和 D_4 的维数估计, 其中模拟了 $N = 10^4$ 个点, 选择了 $k = 100$ 的自举样本, 每个样本包含 $n = 10^4$ 个点间距. 对于 $\varepsilon = 0$ 和 $\varepsilon = 0.3$, 用式子 (8.8) 计算了维数估计. 在水平尺度上, 用 $\overline{y}_{(m)}$ 将结果用点描出来, 这些描出的点是第 m 阶统计 $y_{(m)}$ 的 k 个自举抽样的平均.

用 $\varepsilon = 0.3$ 的一个值做内区域 0.4 的宽度, 从图 9.1 中能看出, 在点间距靠近内区域的宽度之前, 维数估计是 1, 然后比非纠正线下降更快.

考虑 d 维单位立方体中的 N 个一致分布点, 则式子 (9.1) 中给出的 $\mathcal{A}_\varepsilon$ 中点的期望数是 $N(1-2\varepsilon)^d$. 结果, 当 d 增加时, 限制区域中点的期望数以幂指数下降. 于是, 这种方法在某种程度上浪费了数据.

上述方法的另一个问题是, 抽样边界不能很清楚地定义, 也不直接, 就像上面的例子一样. 发生比规定等级更高的地震的地震网中, 能以很高的概率检测到地震, 但地震网的边界经常是不规则的. 它不仅依赖网的密度, 而且依赖不同地区的地理

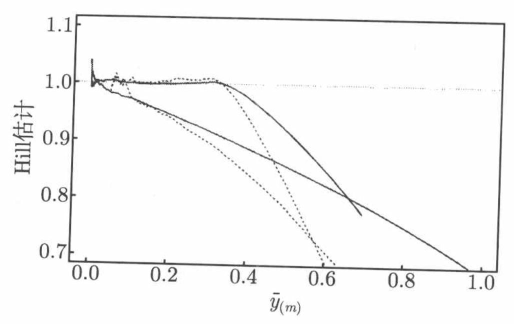


图 9.1 一致随机变量：边界的影响 (Hill 估计)

通过排除一个 $\varepsilon = 0.3$ 的分数来对边界影响的纠正. 实线表示 D_2 的估计, 上面的一条进行了边界纠正. 类似地, 虚线表示 D_4 的估计 $(k = 100, n = 10000, N = 10000)$

特征, 这些特征也影响到能否检测到地震.

另一个方法不是那么浪费数据, 对 $q = 2$ 的情形, 可以利用 D. Stoyan 和 H. Stoyan(1994, §5.5) 给出的 Grassberger-Procaccia 算法 (§8.2). 在 Grassberger-Procaccia 程序中, 统计了一个给定点的 δ 距离内所有点的数目. 如果给定点靠近强加的边界, 所统计的点的数目就将太小. D. Stoyan 和 H. Stoyan (1994) 建议, 用给定点所在球的体积除以与所分析的区域相交的球的体积来改变计数.

Dvořák 和 Klaschka(1990) 建议另外一种纠正, 基于一致分布. 对 $q = 2$, 在 d 维单位立方体中, 点的相关积分是 $F_Y(y) = y^d(2-y)^d$. 他们建议修改 Grassberger-Procaccia 算法 (1983a), 通过计算 $\log F_Y(y)$ 或 $\log[y(2-y)]$ 的斜率, 能将点变换到单位立方体中.

9.2.3 例子——双峰 β 分布

β 分布对于边界的影响提供了一个有趣的反例. 当 $\alpha = \beta = 1$ 时, β 就是一致分布. 当 $\alpha = \beta$ 且都小于 1 时, 分布是双峰的, 并且当 α 和 β 递减时, 质量向区间 $[0, 1]$ 的两端集中. 由于 β 分布的密度是连续的, 于是任给 $q \geqslant 1$, 都有 $D_q = 1$; 然而当 $\alpha = \beta$ 且足够小时, 在 $q = 2$ 的情形, 对于 y 的几乎所有取值, $F_Y(y)$ 能被指数为 2α 的幂指数逼近. 这个例子将在 §10.3.2 中进一步讨论.

9.3 四舍五入的影响

如果数据库中存储的数字只限于表示一个数, 那大部分的数据都被四舍五入了. 四舍五入的程度通常在某种方式上等同于观测的精度. 如果数据四舍五入比较严重, 在维数估计时就有明显的影响, 有下面两个方面的问题. 首先, 抽样的点对会出现很多零间距, 当对点间距取对数时, 那些估计就会出问题; 其次, 有一个调整的

影响, 不应该将它与第 7 章中所讨论的缺项行为相混淆.

处理零差的一种可能的方法是, 利用采样截断的指数分布来修改 Hill 估计.

9.3.1 截断的指数逼近

考虑 $[0,1]$ 内的 n 个点间距 $y_1,\cdots,y_n$ 的样本, 概率分布为 $F_Y(y)=y^\theta$. 假定这些观测值最终调整为 $0,s,2s,\cdots$, 其中 $0<s<1$, 并且 h 调整为 0, $n-h$ 不为 0. 用 Y_i' 表示随机变量 Y_i 调整过的值. 假定这种调整在非零值的分布上不会有太大的影响, 则 $W_i'=-\log Y_i'$ 将会有一个逼近的截断指数分布. 为了记号上方便, 用 $y_1,\cdots,y_{n-h}$ 表示非零的样本值, 用 $y_{n-h+1},\cdots,y_n$ 表示零样本值.

令 $\dfrac{t}{2}=-\log\dfrac{s}{2}$, 则 $\Pr\{Y_i'=0\}=\Pr\left\{Y_i<\dfrac{s}{2}\right\}=\Pr\left\{W_i>\dfrac{t}{2}\right\}=\exp\left(-\dfrac{\theta t}{2}\right)$, 于是用截断的指数分布, log-似然估计为

$$\begin{aligned}L(\theta;w_1,\cdots,w_n)&=\log\left[\binom{n}{h}(e^{-\frac{\theta t}{2}})^h(1-e^{-\frac{\theta t}{2}})^{n-h}\prod_{i=1}^{n-h}\frac{\theta e^{-\theta w_i}}{1-e^{-\frac{\theta t}{2}}}\right]\\&=\log\left[\binom{n}{h}e^{-\frac{\theta ht}{2}}\prod_{i=1}^{n-h}(\theta e^{-\theta w_i})\right].\end{aligned}$$

θ 的极大似然估计为

$$\hat{\theta}=\frac{n-h}{\dfrac{ht}{2}+\displaystyle\sum_{i=1}^{n-h}w_i}. \tag{9.2}$$

9.3.2 修正的 Hill 估计

利用式子 (9.2), §8.5 中的自举抽样程序修正为如下: 式子 (8.7) 变为

$$z_{i,m}=\frac{-1}{m-h}\left[h\log\left(\frac{s}{2y_{(m)}}\right)+\sum_{\substack{j=1\\y_{(j)}\neq 0}}^{m-1}\log\left(\frac{y_{(j)}}{y_{(m)}}\right)\right],$$

式子 (8.8) 还是

$$\hat{\theta}_m=\frac{k}{\displaystyle\sum_{i=1}^{k}z_{i,m}},$$

其中 k 是抽样的数目, h 是抽样的零差的数目.

9.3.3 进一步的问题

前面的方案并不完全正确, 因为在实践中, 四舍五入在更早的阶段, 即在单个的点的位置 X_i $(i=1,\cdots,N)$ 处已经实施了. 令 X_i 是原始点的位置, X_i' 是对应的四舍五入值. 假定随机变量 X_i' 取可能的离散的值 $x_0, x_0+s, x_0+2s, \cdots$, 其中 $s>0$. 令 $Y'=|X_1'-X_2'|$, $Y=|X_1-X_2|$, 则有如下结论:

1. $Y'=0$ 当且仅当 X_1' 和 X_2' 来自于同样的四舍五入状态. 进一步地, 由 $Y'=0$ 导出 $Y\in(0,s)$, 然而由 $Y\in(0,s)$ 不能导出 $Y'=0$.

2. $Y'=s$ 当且仅当 X_1' 和 X_2' 来自于临近的状态. 进一步地, 由 $Y'=s$ 导出 $Y\in(0,2s)$, 然而由 $Y\in(0,2s)$ 不能导出 $Y'=s$.

3. $Y'=2s$ 当且仅当 X_1' 和 X_2' 来自于轮流的状态. 进一步地, 由 $Y'=2s$ 导出 $Y\in(s,3s)$, 然而由 $Y\in(s,3s)$ 不能导出 $Y'=2s$.

4. 等等.

于是四舍五入的差异代表了 Y 在取值上重叠的区间, Y' 的前两种可能都能表示非常小的点间距. 在截断的指数分布中, 用 $Y'=0$ 作为截断的尾, s 作为截断的点, 将很少出现零点, 因为许多小于 s 的值能被包含在 $Y'=s$ 的状态中. 这会导致 θ 的估计太大.

用式子 (9.2) 给出的估计, 对于 $Y'=0$, 能用一个更小的截断点 $\frac{s}{2}$. 这能部分地弥补包含在 $Y'=s$ 中的那些小差异. 用这样的一个截断值, 可以将 Y' 的非零值看做是代表相同长度的非重叠区间.

9.3.4 例子

令 X_1, X_2 是从区间 $[0,1]$ 上的连续一致分布中抽样出来的独立同分布随机变量. 在四舍五入到 p 个小数位后, X_1' 可以取 10^p+1 个可能的离散状态, 概率函数为

$$\Pr\{X_1'=x\}=\begin{cases}\dfrac{1}{2\times 10^p}, & x=0,1,\\[2mm] \dfrac{1}{10^p}, & x\neq 0,1.\end{cases}$$

注意, 并不要求 p 是一个整数, 只要求 10^p+1 是一个正整数就可以了. 现在令 $Y'=|X_1'-X_2'|$, 则可以证明:

$$\Pr\{Y'=y\}=\begin{cases}\dfrac{2\times 10^p-1}{2\times 10^{2p}}, & y=0,\\[2mm] \dfrac{2-2y}{10^p}, & y=\dfrac{1}{10^p},\dfrac{2}{10^p},\cdots,\dfrac{10^p-1}{10^p},\\[2mm] \dfrac{1}{2\times 10^{2p}}, & y=1.\end{cases} \tag{9.3}$$

因此, 当 $\frac{1}{10^p} \leqslant y \leqslant \frac{10^p-1}{10^p}$ 时, Y' 的分布就跟连续且无四舍五入时的三角分布相近. 首先, 对 $y=0,1$, 好像会缺少很多的点差. 但情况不是这样的, 因为 $Y'=0,1$ 表示的区间的宽度是其他状态表示的区间宽度的一半.

于是, 如果将 $Y'=0$ 看做是表示 $Y \in (0, \frac{s}{2})$ 的那些值, 将 $Y'=s$ 看做是表示 $Y \in (\frac{s}{2}, \frac{3s}{2})$ 的那些值, 等等, 则 Y' 的概率分布给出了 Y 的一个很好的逼近. 在某种意义上, Y' 的分布不仅是 Y 的一个离散化的版本, 而且也是一个光滑化的版本.

9.3.5　例子

图 9.2 展示的是四舍五入的影响对一致随机变量的模拟. 模拟了区间 [0,1] 上的 $N=10,000$ 个标量值的一致随机变量, 每一个值都四舍五入到 p 的小数位, 其中 $\log_{10} p = 0.05$, 即每个元被四舍五入到 $0, 0.05, 0.10, \cdots, 1.00$ 这些状态. D_2 的自举抽样 Hill 估计用点描出来了 (实线), 虚线表示没有四舍五入时的维数估计.

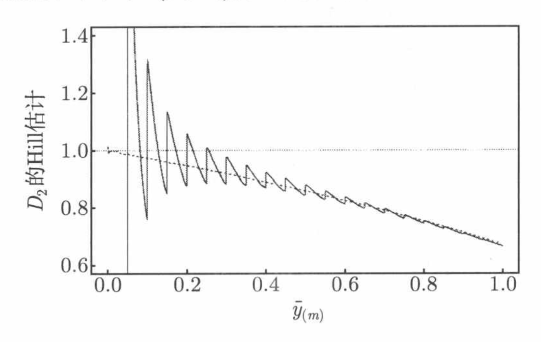


图 9.2　一致随机变量: 四舍五入的影响 (D_2 的 Hill 估计)

$q=2$ 时的 Hill 估计, 此时抽样点在单位区间上一致分布, 但被四舍五入到 $0, 0.05, 0.1, 0.15\cdots, 1$. 水平的点划线仅做参考, 虚线是未做四舍五入时的 Hill 估计 ($k=100, n=10,000, N=10,000, \varepsilon=0$)

调整现象是由严重的四舍五入所引起, 这不会与 §7.4 中的缺陷影响相混淆. 四舍五入时的调整在 $\overline{y}_{(m)}$ 坐标方向有固定的常数周期, 显示出一个锯齿形的影响, 当 $\overline{y}_{(m)}$ 递增时, 振幅递减. 它是由式子 (8.4) 分母中 $y_{(m)}$ 的增量所引起的. 当 $y_{(m)}=0.05$ 时, 阶的统计值都是 0 或 0.05. 于是, 当 $y_{(m)}$ 增加到 0.10 时, 就有一个相当大的影响, 因为对于几乎所有的 $k<m$, $y_{(k)}$ 都是 0 或 0.05, 因而有一个竖直的跳跃. 当 $y_{(m)}$ 递增时, $y_{(k)}$ 的更多的值是 0, 1, 于是水平递减, 直到 $y_{(m)}=0.15$. 重复循环下去.

9.4　噪声的影响

首先区分两种噪声: 观测的误差和系统噪声. 当过程 x_i 被记录为 $x_i+\varepsilon_i$ 时,

就出现了观测误差. 系统噪声对过程来说是内在的, 使得随后的及时观测在功能上依赖先前插入到过程中的噪声, 例如, $x(t_i) = T_\xi[x(t_{i-1})] + \varepsilon_i$. 这一节主要对观测噪声感兴趣.

数据观测上的误差是最主要的一种偏差形式. 令观测区域为 $\chi \subseteq \mathbb{R}^d$, 令测度 μ 支撑在 χ 的一个维数可能低于 d 的子集上. 观测误差的增加往往模糊了任何好的结构, 或填满了整个空间, 使得当点间距递减到零时, 维数的估计会递增到 d.

9.4.1 Cantor 测度

考虑例 8.6.2 中生成的 Cantor 测度, 现在增加白噪声, 结果在图 9.3 中用点描出来了. 当白噪声的标准差增加时, 好的结构渐渐丧失. 于是, 对于更小的点间距, 噪声起主要作用, 维数的估计刻画的是噪声, 而不是 Cantor 测度. 也可以看到, 噪声的影响比四舍五入的影响更严重.

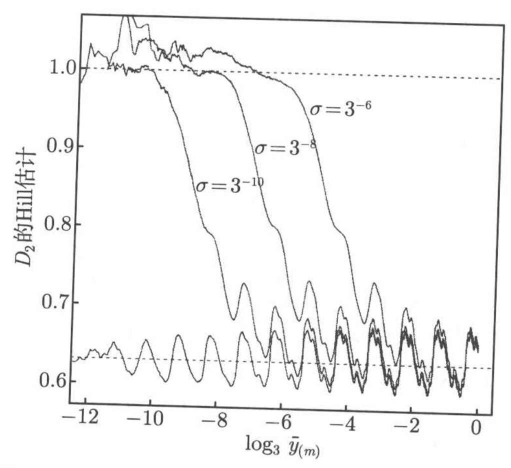


图 9.3 Cantor 测度加白噪声 (D_2 的 Hill 估计)

$p_0 = 0.5$ 时加白噪声的 Cantor 测度的 D_2 估计. 下面的线表示纯 Cantor 测度的估计. 上面的线表示加白噪声的 Cantor 测度的估计, 标准差分别为 $3^{-10}, 3^{-8}, 3^{-6}$. 在 $\log_3 2$ 和 1 处的水平虚线分别表示给定的 Cantor 测度和白噪声的 D_2 的值 ($k = 100, n = 100,000, N = 100,000$)

对数据中噪声的问题, 以及它在维数估计上的影响, Smith(1992b) 也作了讨论.

第10章 维数估计的应用

10.1 引 言

在这一章, 介绍维数估计的一些应用, 这里我们利用这些技巧所分析的数据来自于数学和统计模型的模拟, 而不是真实的数据. 这些数据至少有部分性质是我们已经理解的, 这样有利于评估所用的方法, 以及解释真实情况下出现的结果.

§10.2 包含两个有待考虑的问题. 第一个是给定 $\theta(q)$ 的点态估计, 如何估计重分形谱 $f(y)$? 第二个问题与缺项有关. 在第 8 章所给出的 Cantor 测度的例子中, 缺项循环扩展到点间距 y 的整个取值范围内, 并且没有边界效应. 但对于一般的自相似测度而言, 这一点并不一定总是成立. 下面将给出一个例子, 其中缺项和边界效应同时出现. 在这个给出的例子中, 它们都是内在偏差.

具有地震数据特征的空间点模式的例子将在 §10.3 中讨论. 其中举出的一个例子, 幂律比例关系在点间距取值的大部分范围内成立, 但其指数并不是 Rényi 维数. 其他的例子展示了 Moran 级联过程的某些特征.

在 §10.4, 给出了两个动力系统的例子, 并且估计了它们的 Rényi 维数. 维数的计算建立在整个相空间可观测这个假设上. 但在很多情况, 整个相空间是不可观测的. 此时, 将构造一个重构相空间, 它具有与原来的相空间相同的分形几何性质. 我们会简要的论述这些方法.

给定一组观测时间序列, 一个有趣的问题是, 这些数据是否由一个有限维动力系统生成, 以及这个系统是否是随机的. 由此引出的问题是, 一个决定性过程和一个随机过程之间存在何种差异, 以及如何在经验观察的基础上区别这些差异. 这些问题将在 §10.5 中探讨.

自相似随机过程同样满足各种幂律关系, 特别是自相关函数, 以及增量大小尺度随时间的分布方式. 从 Lamperti 的文章 (1962) 开始, 许多文献详细研究了自相似随机过程. 然而, 自相似随机过程从一开始就被定义为单重分形的 (即只有一个幂律指标). 最近, 对这些过程的研究扩展到了满足多重比例性质的“重分形”随机过程, 正如对重分形测度的研究. 本书将在 §10.6 中介绍自相似随机过程及重分形随机过程.

10.2 进一步的估计和诠释

本节的第一个例子说明, 在对自相似集上的一个分布 (多项分布测度) 作取样

观测计算时，得到的维数估计可能既有缺项又有边界效应. 在前一章，研究了 Cantor 测度，它只能引出边界效应. 关于测度如何构造的信息能从边界效应的表现形式中导出.

第二个例子是关于重分形谱的一个估计，其计算是在估计 Rényi 维数的基础上，用 Legendre 变换得到的.

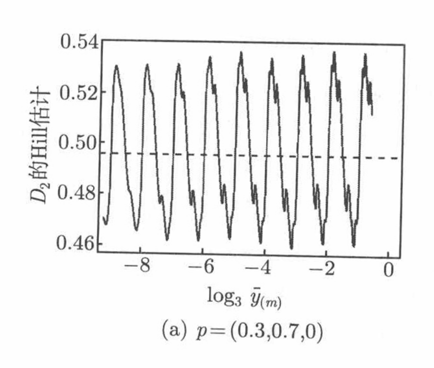
10.2.1 例子——多项分布测度

考虑支撑在压缩比都相等的自相似集上的测度，其构造方法是按一个一致的权重，将质量逐级分配给自相似集的每个相似基本块 (例：具有相同压缩比的 Moran 级联过程). 该测度给出一个缺项的概率分布 $F_Y(y)$，其中 Y 为 q 阶点间距. 对于 Cantor 测度，缺项是内在偏差的唯一表现 (见图 7.4)，并且图 7.2 所示的一致分布没有整体边界效应，即 $\Psi(y)$ 在 y 的整个取值范围内逐步减少. 边界效应的缺失是由“间隔”的位置引起的. 若多项分布测度没有定理 5.5.2 定义的“间隔”，则既不缺项，也没有整体边界效应.

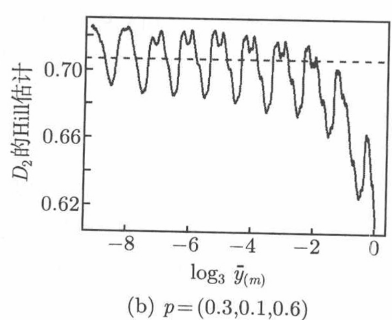
图 10.1 画出了从各种多项分布测度所得到的模拟样本的维数估计 ($N = 100,000$)，测度的参数标在每幅图的上方，其中 $p = (p_0, p_1, \cdots, p_{b-1})$. 每个选点都是通过一个 b-进序列的模拟得到. 序列中每位数取 ω 的概率为 p_ω，其中 $\omega \in \{0, 1, \cdots, b-1\}$. 然后转化为一个 0,1 之间的十进小数. 从 $n = 100,000$ 个点间距中取 100 个自举抽样 ($k = 100$) 进行模拟.

当“间隔”出现在开始或结尾，即 $p_0 = 0$ 或 $p_{d-1} = 0$，则最后一个缺项循环似乎是不完全的，见图 10.1(a), (e). 当没有“间隔”时，$F_y(y)$ 是缺项的，并且有前述的边界效应，见图 10.1(b). 当有“间隔”时，似乎没有边界效应，见图 10.1(f). 若仅出现某些“间隔”，边界效应会以不同的强度出现，试比较图 10.1(c), (d). 注意到所有这些例子都有相同的 Rényi 维数 $\widetilde{D}_2$，可能 D_2 也相同，但每个相对应的 $F_Y(y)$ 都是不同的. 定理 5.5.4 断言 $\widetilde{D}_2 = D_2$，只对有间隔的多项分布测度成立.

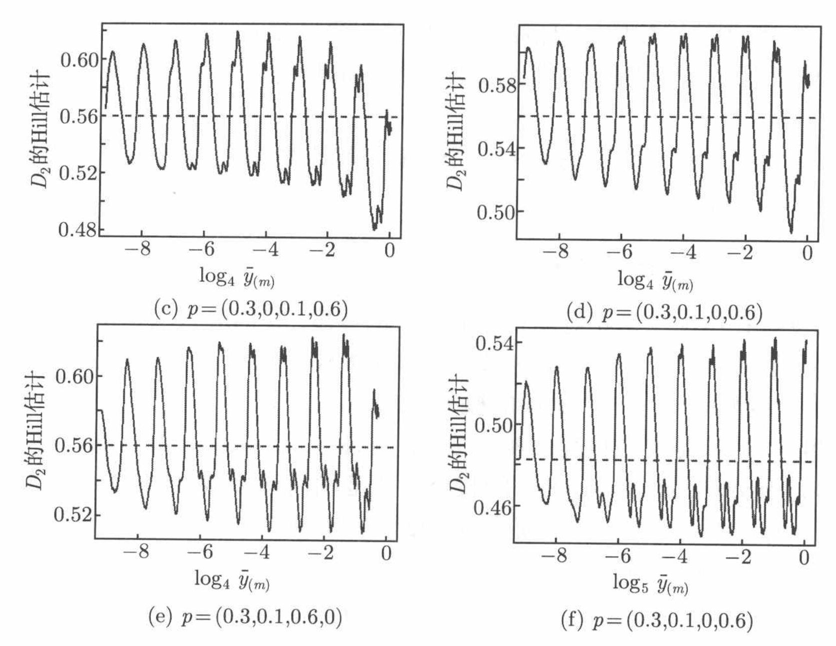
注意到 $F_Y(y)$ 显示的边界效应为一个周期部分，其周期在 b 为底的对数尺度



(a) $p=(0.3,0.7,0)$



(b) $p=(0.3,0.1,0.6)$



(c) $p=(0.3,0,0.1,0.6)$ (d) $p=(0.3,0.1,0,0.6)$

(e) $p=(0.3,0.1,0.6,0)$ (f) $p=(0.3,0.1,0,0.6)$

图 10.1 有或没有间隔的多项分布测度

对于多种有或没有“间隔”的多项式分布测度的 D_2 的估计, 其中 $p=(p_0,p_1,\cdots,p_{b-1})$. 水平间断线为

$$\widetilde{D}_2=-\log_b\sum_{\omega=0}^{b-1}p_\omega^2,\ (k=100,\ n=100,000,\ N=100,000)$$

上为 1. 在每个多项分布测度的构造中, 相似比都为 $1/b$. 一般而言, 自相似集可以有不同的压缩比. 当一个多项分布测度支撑在这样的自相似集上, 例如 Moran 级联测度, 会出现不同的缺项循环相互重叠的情况. 同样的, 在对 D_2 的维数估计中, 这些缺项循环表现为噪声的形式, 见例 10.3.1.

10.2.2 例子——Cantor 测度

再次考虑 $p_0=0.2$ 的情形, 其维数估计见图 8.6, 修正的 Rényi 维数估计见表 8.1. 把这些记为 $\widehat{D}_q$, 对于 $q=2,\cdots,8$, 并将相应的 $\theta(q)$ 的点估计记为 $\widehat{\theta}(q)=(q-1)\widehat{D}_q$. 这里, 试图利用这些点估计部分地重建方程 (2.9) 所给出的重分形谱 $f(y)$. 为此, 对于 $q=2,\cdots,8$, 将根据点估计 $\widehat{\theta}(q)$, 选取合适的函数. 记这样的函数为 $\widehat{\theta(q)}$. 然后, 试图取 $\widehat{\theta(q)}$ 的 Legendre 变换, 作为 $f(y)$ 的估计函数, 记为 $\widehat{f(y)}$.

$\theta(q)$ 的 Legendre 变换为 $f(y)=\inf_q\{qy-\theta(q)\}$. 可以表示为 $f(y(q))=qy(q)-\theta(q)$, 其中 $y(q)$ 为 $\theta(q)$ 的导数. 因为 $\theta(q)$ 为一个重比例化累积生成函数, 它的导函数具有图 3.3 中所示的图像. 即 $\theta(q)$ 随着 q 的增长, 单调下降到一水平渐近线, 并且随着 q 的减少, 单调上升到另一水平渐近线 (见图 3.4). 图 3.3 中导函数的形

状与对数函数很相似. 设下渐近线为 α_1, 上渐近线为 α_2. 修改对数函数可以给出 $y(q)$ 的估计函数的某种参数形式:

$$\widehat{y(q)} = \frac{\alpha_1 e^{-q}}{e^{-q}+1} + \frac{\alpha_2 e^{q}}{e^{q}+1}.$$

然后积分可得 $\widehat{\theta(q)}$ 的函数式. 添加适当的常数使得 $\widehat{\theta(1)} = 0$, 得到

$$\widehat{\theta(q)} = \alpha_1 \log(e^{-1}+1) - \alpha_2 \log(e^{1}+1) - \alpha_1 \log(e^{-q}+1) + \alpha_2 \log(e^{q}+1).$$

如果有 $\theta(q)$ 更多的点估计, 可以增加附加常数以修正函数逼近两条渐近线的速度. 就现有的数据, 一个两参数的模型已经足够了.

非线性的最小平方法可以用来估计 α_1 和 α_2, 得到的估计为 $\widehat{\alpha}_1 = 0.964140$ 及 $\widehat{\alpha}_2 = 0.195159$. 利用方程 (3.15) 和 (3.16), 可以解出 α_1 和 α_2 的真实值分别为 1.464974 和 0.203114. 对 α_1 的估计糟糕的原因是没有 $q \leqslant 0$ 的数据.

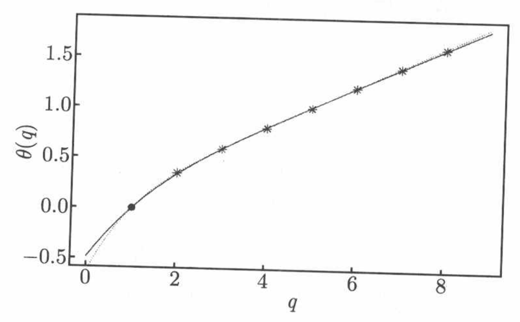


图 10.2 当 $p_0 = 0.2$ 时, Cantor 测度的 $\theta(q)$ 的估计

实线为 $p_0 = 0.2$ 时 Cantor 测度的 $\widehat{\theta(q)}$. 星点表示 $q = 2, \cdots, 8$ 时对 $\widehat{\theta}(q)$ 的估计, 实心点表示 $\theta(1) = 0$ 这个固定点. 点线表示方程 (3.8) 给出的函数 $\widetilde{\theta}(q)$

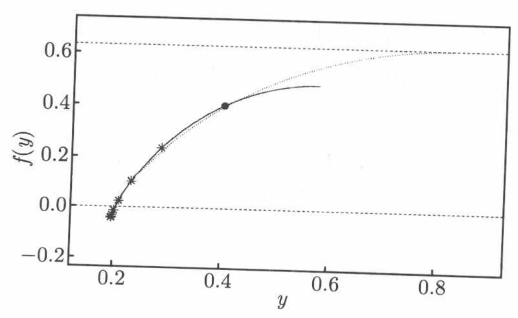


图 10.3 当 $p_0 = 0.2$ 时, Cantor 测度的 $f(y)$ 的估计

实线为 $p_0 = 0.2$ 时 Cantor 测度的 $\widehat{f(y)}$). 实心点表示在 $\widehat{y(1)}$ 处的固定点, 星点标注 $\widehat{y(2)}, \cdots, \widehat{y(8)}$ (从上往下). 点线表示由方程 (3.6) 给出的多重分形谱. 上面那条间断线表示 $f(y(0)) = D_0$

对于 $0 < q < 9$, 图 10.2 画出了函数 $\widehat{\theta(q)}$ 的图像, 并且与 $\widetilde{\theta}(q)$ 的已知数值相比较. 可以估算出重分形谱为

$$\widehat{f(\widehat{y(q)})} = q\widehat{y(q)} - \widehat{\theta(q)}.$$

图像及与 $\widetilde{f}(y)$ 已知值的比较见图 10.3. 注意, 与图 3.2 一样, $f(y)$ 是正的. $\widehat{f(y)}$ 的负值是由于当 q 较大时, 点估计 $\widehat{\theta}(q)$ 较小, 见图 10.2.

10.3 空间与时间点模式

涉及指数估计的各种方法, 如关联维数, 经常被用于点过程分析来刻画空间点模式的聚集程度. 例如, Kagan 和 Knopoff (1980) 通过计算他们定义的两点关联函数, 来刻画地震集群. 在本质上, 这与关联维数 D_2 是一致的. Kagan (1981a,b) 将之扩展到了三点和四点关联函数. 它们与 D_3 和 D_4 并不完全相同. 三点关联函数基于三点所成的三角形的面积, 四点函数则基于四面体的体积. 这些三角形和四面体的幂律数被确定为与它们的大小 δ 相关的一个函数. 然后, Kagan 将之应用到简单 Poisson 分布中, 利用这些函数的数值将所得到的估计标准化.

幂律估计并不一定有一个“分形”的解释. 一个集合的分形维数刻画了当覆盖所用集合的大小趋于无限小时, 覆盖该集合所需的集合的个数的幂律增长. 当集合为自相似集时, 幂律在所有的尺度上都成立. 在其他情况下, 幂律可能只在一个适当的尺度范围内成立, 当尺度过大或过小时, 都不再成立. 将在例 10.3.2 中看到这一点, 其中, D_2 基于从 Beta 分布中得到的样点对 X_1 和 X_2. 在这种情况下, D_2 为 1, 虽然可以选择 Beta 分布的参数, 使得 $F_Y(y)$ 在 y 的几乎整个取值范围内满足幂律, 并且幂律的指数可显著小于 1, 其中 $Y = |X_1 - X_2|$.

幂律关系不能扩展到无限小尺度的原因可能是由于数据的缺乏, 尽管幂律可能是无限小尺度上一个相当明智的模式假设. 另一种可能是, 在越来越小的尺度上, 幂律可能不是一个合理的模式假设. 在计算“维数”时, 需要明白这些数字的意义是一个真实的维数, 即表明幂律在无限精细的尺度上都成立, 还是仅仅表明幂律只在一个有限的尺度范围内成立. 如果是后者, 这些指数实际上与点间距的概率分布有关. 在这一节, 将进一步讨论这些问题中与空间点模式和点过程相关的问题.

10.3.1 例子——Moran 级联过程

空间点模式及所带的一个概率分布可由如下的 Moran 级联过程给出. 考虑例 6.2.3 中的情形. 初始圆周半径为 1, 圆心在 $(0,0)$. 在这个圆中, 有四个小圆, 圆心分别为 $(V_0, W_0) = \left(\frac{1}{2}, 0\right)$, $(V_1, W_1) = \left(0, -\frac{2}{3}\right)$, $(V_2, W_2) = \left(-\frac{1}{2}, 0\right)$ 和 $(V_3, W_3) =$

$\left(0,\dfrac{2}{3}\right)$, 相似比分别为 $t_0,\cdots,t_3$. 这里 $b=4$, $\Omega=\{0,1,2,3\}$. 给定一个随机序列 $(\omega_1,\omega_2,\cdots)\in\Omega^\infty$, 其中 $\omega_i\in\Omega$, 利用方程 (6.1) 中的编码映射, 可以确定模拟点的坐标. 数字 $j\in\Omega$ 出现的概率为 p_j, 这里 $\sum\limits_{j\in\Omega}p_j=1$. 对于我们的模拟来说, 一个长度为 $n=40$ 的序列就足够了, 模拟点的坐标由如下的式子给出:

$$x=V_{\omega_1}+\sum_{k=1}^{n-1}\left(V_{\omega_{k+1}}\prod_{i=1}^{k}t_{\omega_i}\right),$$

及

$$y=W_{\omega_1}+\sum_{k=1}^{n-1}\left(W_{\omega_{k+1}}\prod_{i=1}^{k}t_{\omega_i}\right).$$

图 10.4 显示了两个具有 1000 个点的模拟, 它们具有不同的 p_j. 在图 10.4(a) 中, 所有的 $p_j=0.25$, 从而每个点族有同样多的点. 上下两个点族具有更高的密度是因为它们的相似比较小. 第二个过程在左右两个点族上分配了更高的概率, 从而看上去更多的点都集中在中间的平行线上. 这些点可以表示一个平面物体的微小应力破裂的初始点. 很多这样的微小破裂组成的直线形成了一个由多个小的平行断层组成的大断层.

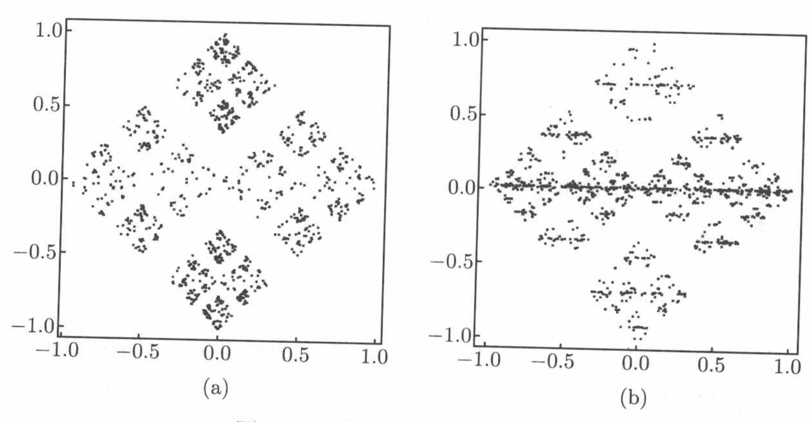


图 10.4　模拟 Moran 级联过程

模拟 Moran 级联过程 ($N=1000$), 相似比 $t_0=\dfrac{1}{2}$ (右), $t_1=\dfrac{1}{3}$ (下), $t_2=\dfrac{1}{2}$ (左), $t_3=\dfrac{1}{3}$ (上). 图 (a) 中所有 $p_j=0.25$, 图 (b) 中 $p_0=p_2=0.4$, $p_1=p_3=0.1$

图 10.5(a) 显示了图 10.4(a) 所示过程的 D_2 维数估计, 取样大小为 $N=10000$. 注意到这一过程具有不同的相似比, 所以其缺项循环并不显著. 图 10.5(b) 显示了同样过程的维数估计, 但所有的相似比都取 $\dfrac{1}{3}$. 这样就产生了一个明显的周期, 取

以 3 为底的对数即可看出. 注意, 当压缩比都相等时, 边界效应是如何在后来产生的 (缺项循环相互间有效协调).

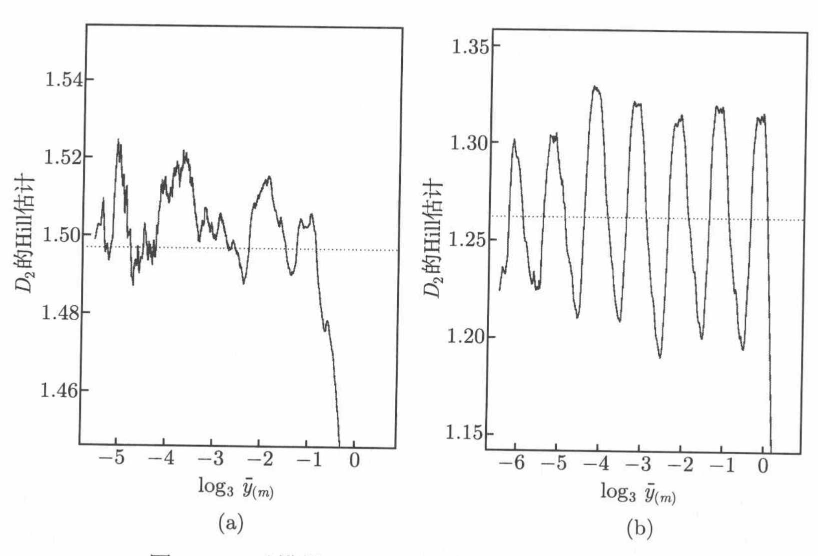


图 10.5　对模拟 Moran 级联过程的 D_2 的估计

对于一个长度为 $N = 10,000$ 的模拟 Moran 级联过程的 D_2 的估计. 两种情况中, $p_j = 0.25$, $j = 0, \cdots, 3$. 水平的点线为 $D_2^\star$. 图 (a) 的过程与图 10.4 (a) 图的过程一样, 其中相似比为 $t_0 = \frac{1}{2}$ (右), $t_1 = \frac{1}{3}$ (下), $t_2 = \frac{1}{2}$ (左), $t_3 = \frac{1}{3}$ (上). 图 (b) 的过程具有相同的相似比, 即 $t_j = \frac{1}{3}$, 当 $j = 0, \cdots, 3$

空间点模式的维数估计有时会被两个较其他点族大很多的点族所决定. 这可以在下例中看出, 其中的幂律出现在对 D_2 的估计中, 尽管在对高阶 Rényi 维数的估计中会得到峰值.

10.3.2　Beta 分布

Beta 分布可以用来给出一个 §9.2 中所讨论的边界效应的有趣的对照. 边界效应是因大的点间距相对稀少所导致的. 如果从满足 $\alpha = \beta < 1$ 的 Beta 分布中取点对, 相比 $\alpha = \beta = 1$ 的情形 (即一致的), 将会得到更多的大的点间距, 这是因为当 α 和 β 减少时, 质量会更多的集中在单位区间的边界上.

进一步, $Y = |X_1 - X_2|$ 的概率分布在 Y 的很多取值范围内, 都越来越接近以 2α 为指数的幂律. 在图 10.6 中可以看到这一点. 当 $\alpha = \beta = 0.5$ 时, D_2 的估计中出现的边界效应相对于 $\alpha = \beta = 1$ 时 (即一致的) 要微弱的多. 对于 $\alpha = \beta$ 取更小的值的情形, 代表 D_2 估计的直线表明, 当 $q = 2$ 时, $F_Y(y)$ 在它的几乎所有取值范围内满足幂律.

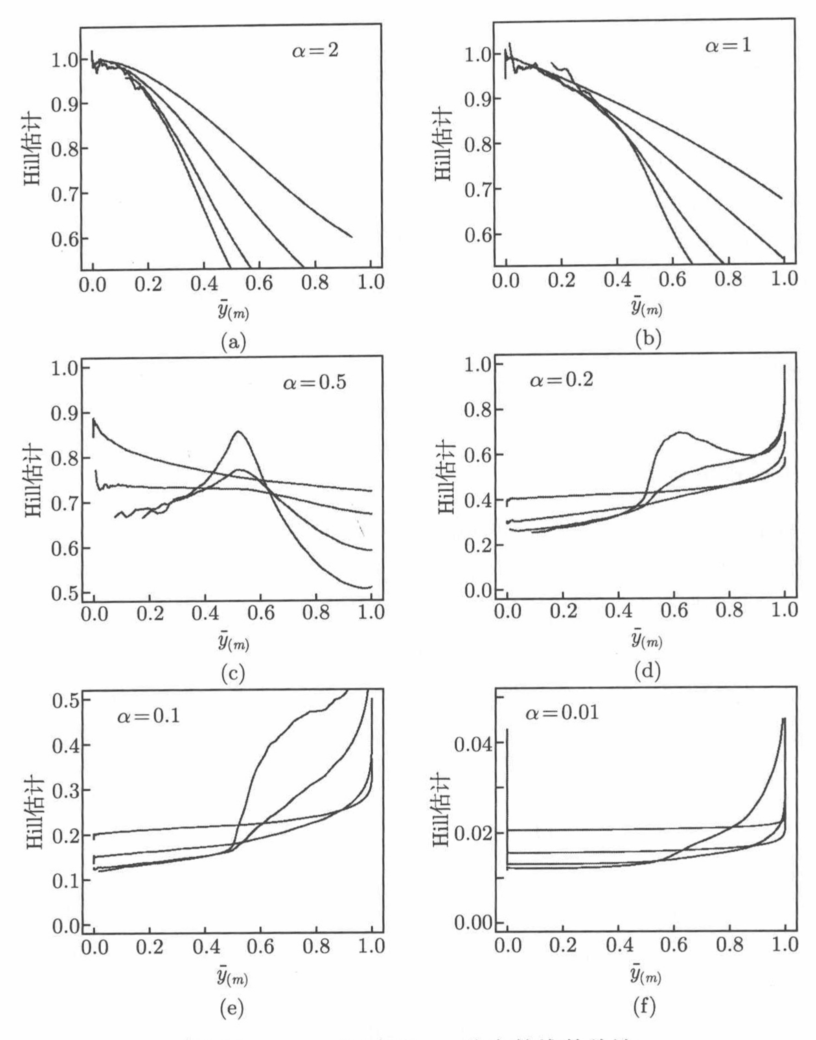


图 10.6 $\alpha = \beta$ 时, Beta 分布的维数估计

从一个 $\alpha = \beta$ 的 Beta 分布取样本点所得到的 D_2, D_3, D_5 和 D_7 的估计. 在图 (a), (b) 中, 从上到下的各条曲线分别表示 D_2, D_3, D_5 和 D_7. 在图 (c)~(f) 中, 当 y 取值较小时, 从上到下的各条曲线分别表示 D_2, D_3, D_5 和 D_7, 但最终最上面的曲线为 D_7

这样的幂律行为可以用关于两个 Beta 随机变量的差的已知结果来验证. 在我们的例子中, $\alpha = \beta$, 从而差的分布关于原点对称. 根据 John, Kota 和 Balakrishnan (1994, 方程 (25.103e)) 的结果, Y 的概率密度函数 $f_Y(y)$ 为

$$f_Y(y) = \frac{y^{2\alpha-1}(1-y)^{2\alpha-1}}{B(\alpha,\alpha)} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} A_{\alpha,i,j} \frac{(1-y)^i(1-y^2)^j}{i!\,j!},$$

其中 $B(\alpha,\beta)$ 为 Beta 函数, $A_{\alpha,i,j}$ 为非负常数. 当 $y>0$ 时, 无穷级数收敛. 将 $(1-y)^{2\alpha-1}$ 表示为 Taylor 级数 $(y<1)$, 密度函数可以表示为一个首项为 $y^{2\alpha-1}$ 的多项式. 积分后得到, 当 $\alpha=\beta$ 充分小时, Y 的概率分布满足以 2α 为指数的幂律. $\alpha=\beta$ 越小, 首项就越逼近真实值. 当 α 和 β 充分小, Beta 分布本质上就是在单位区间端点的两个幂律分布.

高阶 Rényi 维数估计表明 Beta 分布不是多重分形. 对于 D_2, 幂律不成立. 进一步, 单峰随着 q 的增长而增长, 这表明两个点族之间的相互作用是本质的.

之所以介绍上面的例子, 是因为它与地震空间点模式所表现的行为并不是完全不同的, 这里两个密度非常大的点族对维数估计起决定作用. 这将在第 11 章中作进一步讨论.

10.3.3　点过程的相容性

当维数估计是通过对观测点模式的计算得到时, 对估计的解释会有某些不明确的地方. 有限的点模式总是 0 维的. 例如, 对于一个从 Moran 级联模式模拟得到的点模式, Rényi 维数估计与数据取样所用的概率分布有关.

Vere-Joes(1999) 证明, 对于一个观测空间 - 时间点模式的 Rényi 维数估计, 至少有两种不同的解释. 第一种, 固定空间区域, 估计是关于增长的时间段, 第二种, 固定时间段, 让空间区域的大小增长. §11.1.1 会进一步讨论这两种情况.

10.4　动力系统

在这一节中, 将估计 Logistic 映射和 Lorenz 吸引子的 Rényi 维数. 维数估计中的缺项行为清楚地表明了在 Logistic 映射这个例子中的类 Cantor 结构.

一个动力系统往往只有它的一个投影是能被观测的, 通常是一个标度时间序列. 即整个过程不是完全可观测的. 利用观测时间序列, 人们希望能确定无法整体观测的过程的分形特征. 下面将讨论这个问题, 并将其运用到 Lorenz 吸引子中.

10.4.1　Logistic 映射

已经在 §1.3.2 中讨论过 Logistic 映射. 随着 ξ 增长到 $\xi_\infty\approx 3.569945672$, Logistic 映射的性态经历了一系列的分叉点, 见 §1.3.2. 当 $\xi=\xi_\infty$ (见图 1.4), 吸引子集在本质上类似于 Cantor 集.

在这个例子中, 对类 Cantor 情形估计 Rényi 维数 $(q=2,\cdots,5)$, 其中 $\xi=\xi_\infty$. 在我们的模拟中, 用 $\xi=3.569945672$ 作为 ξ_∞ 的近似值. 使用 ξ_∞ 的近似值将导致有趣的行为. 如果模拟一个足够长的序列, 它将重复它的轨迹, 即使周期非常大 (见 Grebogi et al., 1998). 这将导致对于小的点间距, 维数估计值趋于 0. 当模拟序列足

够短, 以致观测不到周期行为, ξ_∞ 的近似值类似于对模拟序列添加一个噪声, 并且对小的点间距, 维数估计值趋于 1.

如果模拟序列开始重复自己, 存在一个合理的可能性, 使得零点间距被取到. 方程 (8.8) 给出的 Hill 估计无法处理零距离. 在实际操作中, 使用 §9.3.2 中给出的修正 Hill 估计, 并且假定机器精度大概为 10^{-16}. 在 §9.3.2 中, 已将 h 设定为这个值.

上述的效应可以在图 10.7 中看出. 取 ξ_∞ 的五个近似值, 用来模拟长度为 $N=100000$ 的序列. 对于每个序列, 计算 D_2 的 Hill 估计用到了 $k=100$ 个自举抽样, 每个取样由 $n=100000$ 个点间距取样组成. 水平轴标示了在以 6 为底的对数尺度上的点间距, 这样做的目的是为了逼近图 1.3 所确定的自相似压缩因子. 当 $\xi=3.568$ 或 3.69 时, 零点间距被取到, 使得对于小的点间距的维数估计为 0. 当 ξ 取其他值, 对于大的点间距, 过程显示出类 Cantor 的行为. 这可以从类缺项行为的发展接近吸引子的 Hausdorff 维数 (加点线) 看出 (Falconer, 1990, 176 页). 对于小的点间距, 维数估计值趋于 1, 即由于使用 ξ_∞ 的近似值, 精细结构变得模糊了.

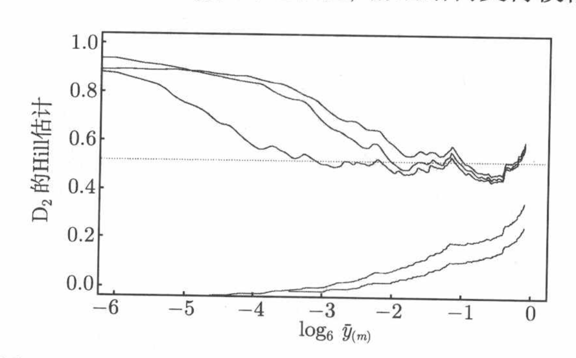


图 10.7 当 ξ 逼近 ξ_∞ 时, 对 Logistic 映射的 D_2 估计

对 Logistic 映射的 D_2 估计, 其中从下到上的曲线分别对应 $\xi=3.568,\ 3.569,\ 3.570,\ 3.571,\ 3.572$ $(k=100,\ n=100000,\ N=100000)$

当取 $\xi=3.569945672\approx\xi_\infty$ 时, 图 10.8 画出了 $D_2,\cdots,D_5$ 的维数估计. 对于 Logistic 映射, 要在一些周期上得到明显的类缺项行为, 需要模拟一个非常长的序列. 为了完成图 10.8 中的计算, 需要模拟一个长度为 $N=300000$ 的序列, 并且用到了 $k=100$ 个自举抽样, 每个由 $n=300000$ 个点间距取样组成. 注意到, 缺项循环在以 6 为底的对数尺度上看上去具有大致的周期性, 与图 1.3 一致. 与图 8.5 和 8.6 相比, 精度较低, 其原因是吸引子集合包含在单位区间的一个很小的部分 (见图 1.4), 并且在模拟中取的是 ξ_∞ 的近似值. 表示 D_2 的维数估计的曲线在 −7.0 到 0.5 之间几乎是水平延伸, 然后趋于 1. 这与使用 ξ_∞ 的近似值所导致的误差也是一致的.

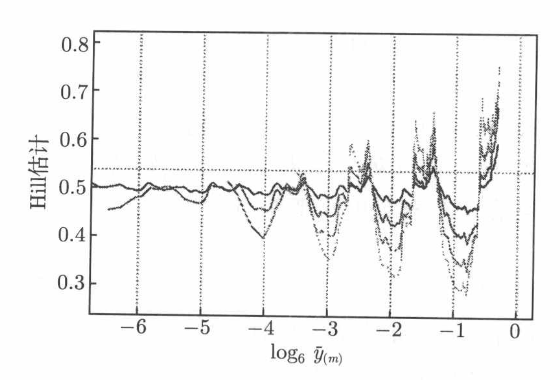


图 10.8 当 $\xi = 3.569945672$ 时, 对 Logistic 映射的维数估计

当 $\xi = 3.569945672$ 时, 对 Logistic 映射的 $D_2, \cdots, D_5$ 的估计. 点线表示吸引子 (Falconer, 1990, 176 页) 的 Hausdorff 维数 $(0.538\cdots)$. 从具有最小峰值和最大极小值的曲线, 到具有最大峰值和最小极小值的曲线, 分别表示 D_2 到 D_5 ($k = 100$, $n = 300000$, $N = 300000$)

对 Logistic 映射的 D_2 估计, 其中从下到上的曲线分别对应 $\xi = 3.568, 3.569, 3.570, 3.571, 3.572$ ($k = 100, n = 100,000, N = 100,000$).

利用图 10.8, 可得到 D_2 的一个粗糙估计为 0.5. 与图 8.6 所示, 取 $p_0 = 0.2$ 的 Cantor 测度相比, D_3, D_4 和 D_5 的估计数值仅有一些小的下降, Logistic 映射的 D_5 比 0.48 略小. 这预示着, 相比取 $p_0 = 0.2$ 的 Cantor 映射的情形 (见例 1.3.1), Logistic 映射的轨道可能更平均的跑到吸引子集合的各个子集.

10.4.2 推广的面包师映射

面包师映射得名于它由一个拉伸变换, 再做一个折叠变换得到, 就像面包师准备生面团时的操作. Falconer (1997, 177 页) 讨论了这个映射. Smith (1992a) 给出了这一映射的推广, 称之为推广的面包师映射, 并且计算了当 $q = 2$ 时, 相应的函数 $\Phi(y)$. 在这种情形, $\Phi(y)$ 具有与图 7.5 所示的 Cantor 测度极为相似的性质.

10.4.3 Lorenz 吸引子

如例 1.3.3 所示在 $\mathbb{R}^3$ 中定义 Lorenz 方程组. 将这个方程组所定义的整个系统设为相空间. 当估计一个可观测动力系统的分形维数时, 常常假设整个过程只有一部分是能被观测的, 即在一个观测空间中观测整个过程. 因此, 假设存在一个从相空间到观测空间的投影, 其中, 我们的观测可能仅仅由标度时间序列组成.

在这个例子中, 假设观测空间就是相空间, 并且观测点在轨道上的位置包含了所有的三个维度. 将在 §10.4.4 中讨论, 当相空间和观测空间不一样时, 会发生什么情况.

在这个例子中, 利用方程 (1.1) (Euler 方法) 所给出的逼近来模拟一个步长为

$h = 0.001$ 的序列. 模拟观测次数 (在 $\mathbb{R}^3$ 中) 为 $N = 2 \times 10^6$. 取 $n = 2 \times 10^5$ 个点间距进行维数估计, 并重复做 $k = 100$ 次自举抽样. 图 10.9 画出了 Hill 估计.

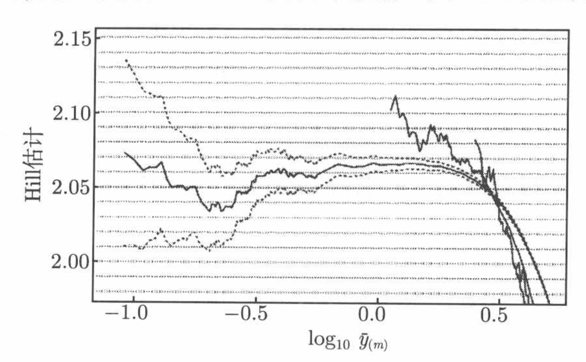


图 10.9 Lorenz 吸引子的维数估计

Lorenz 吸引子的 $D_2, \cdots, D_5$ 估计 (实线). 表示 D_2 的曲线延伸到最左边, 然后依次是 D_3, D_4 和 D_5. 两条间断线表示以 D_2 的估计为中心的区间, 其半宽度为两倍标准误差 ($k = 100$, $n = 200000$, $N = 200000$)

图 10.9 还画出了表示 D_2 维数估计两倍标准误差的区间. 容易看到, 对于越小的点间距, 估计的变化越剧烈, 但同样的, 随着 m 的增长, 自相关性也是明显的; 试比较图 8.1.

同样注意到, D_3 到 D_5 的估计没有扩展到充分小的点间距, 幂律行为对于这样的点间距是显然的. 这是在估计高阶 Rényi 维数时产生的一个问题. 为了使维数估计不被边界影响, 并且落在一个幂律比例关系明显的范围内, 需要一个长的观测序列.

10.4.4 嵌入和重建

到目前为止, 所做的关于动力系统的讨论, 总是假设整个过程在 $\mathcal{X} \subseteq \mathbb{R}^d$ 上是可以被直接观测的, 并且可以表示为一个直接的时间映射

$$x(t_{n+1}) = T_\xi[x(t_n)] = T_\xi^{n+1}[x(t_0)].$$

物理学家将 $\mathcal{X}$ 称为相空间. 目标是刻画概率测度 μ, 它表示在任意时刻轨道落在集合 A 中的概率. 因为过程是决定性的, 并且测度 μ 被假定为关于映射 T_ξ 不变, $\mu(A)$ 等价于 $x(t_0) \in A$ 的概率. 条件 μ 不变等价于时间序列 $\{x(t_n)\colon n = 0, 1, \cdots\}$ 是严格稳定的.

通常, 一个动力系统的相空间 $\mathcal{X}$ 是不能直接观测的, 只能在另一个空间中观测整个过程. 设这个空间为 $\mathcal{Y} \subseteq \mathbb{R}^k$, 其中 $k < d$, 它是 $\mathcal{X}$ 的一个投影, 即

$$\Pi\colon \mathcal{X} \to \mathcal{Y}.$$

称 $\mathcal{Y}$ 为观测空间. 通常 $k=1$, 则我们的观测是一组单变量的时间序列 $\{y(t_n)\colon n=0,1,\cdots\}$. 给出了这些观测后, 能确定概率测度 μ 的重分形特征吗?

重分形特征常常是通过将时间序列嵌入一个 p 维的重建相空间得到. 这通常是用时间延迟的方法来做. 其中, 一组新的 p-维时间序列, 记为 $y^{(p)}(t_n)$, 定义为

$$y^{(p)}(t_n)=(y(t_n),y(t_{n+\ell}),\cdots,y(t_{n+(p-1)\ell})), \tag{10.1}$$

其中, $\ell\in\mathbb{Z}^+$ 为一个延迟参数. 所需的 p 值将在下面讨论. 现在将 $y^{(p)}(t_n)$ 重新表示为

$$y^{(p)}(t_n)=\left(\Pi\circ T_\xi^n[x(t_0)],\Pi\circ T_\xi^{n+\ell}[x(t_0)],\cdots,\Pi\circ T_\xi^{n+(p-1)\ell}[x(t_0)]\right).$$

得到一个复合映射, 记为 Ψ_p

$$\Psi_p\colon\mathcal{X}\to\mathcal{Y}^p, \tag{10.2}$$

其中, $\mathcal{Y}^p$ 为 $\mathcal{Y}$ 的乘积空间. 若 $\mathcal{X}$ 的重分形性质在重建相空间 $\mathcal{Y}^p$ 中保持不变, 则称 Ψ_p 为嵌入映射. 通常, 嵌入概念要求微分结构也保持不变, 但在这里这一点并不是必需的.

如方程 (10.1) 所示的时间延迟的方法, 是基于 Takens (1981) 的一个结果, 这个结果基于 Whitney (1936) 的嵌入定理. 这样的时间延迟重建是 Packard 等 (1980) 首先提出的. Sauer 等 (1991) 证明, 若方程 (10.1) 中的 p 大于 $\mathcal{X}$ 中吸引子的盒维数的两倍, 则 Ψ_p 为一个嵌入映射.

关于嵌入概念的进一步讨论可以在 Whitney (1936), Sauer 等 (1991), Cutler (1997), Cheng 和 Tong (1994), Isham (1993) 中找到. Casdagli 等 (1991) 在有噪声的情况下, 讨论了状态矢量空间的重建, Fraser 和 Swinney (1996) 讨论了对于重建的最优时间延迟长度. 进一步的讨论也可以参考 Cutler 和 Kaplan (1997), 其中包括了一系列的评论文章, 作者既有数学家也有物理学家. Abarbanel 等 (1993) 的一篇关于观测到的混沌数据的分析的文章, 以及 Abarbanel (1995) 的一本书也包含了关于相空间重建的极好的章节.

10.4.5　例子——Lorenz 吸引子 (续)

在这个例子中, 仅仅假设 Lorenz 吸引子的 $x_1(t_n)$ 部分可以被观测, 即 $y(t_n)=x_1(t_n)$, 其中, $x_1(t_n)$ 由方程 (1.2) 定义. 利用这个标度时间序列, 可以用时间延迟方法, 通过 $y^{(p)}(t_n)$, 得到一个到重建相空间中的嵌入. 如例 10.4.3 所分析的, 它具有与 $\mathcal{X}$ 一样的分形性质.

我们知道 p 一定要大于吸引子盒维数的两倍. 但仍然需要确定延迟参数 ℓ. 选择一个最佳值并不容易. 观测一个具有有趣的自相关性的系统. 然而, 如果 ℓ 过小, 那么只能从持续的观测中得到很少的信息. 例如, 如果以间隔为 1 秒的时间序列来

对惠灵顿的温度做观测, 则所得结果可能含有相当多的多余信息. 相反, 如果 ℓ 过大, 将难以得到相邻两个观测值的关系, 并且, 这些观测值可能在本质上是互不相关的. 图 10.10 中, 对于各种不同的 ℓ 的取值, 对比画出了标度时间序列 $y(t_n)$ 和 $y(t_{n+\ell})$. 当 ℓ 的取值在 10 到 20 之间时, 吸引子的结构呈现出来, 试比较图 1.5. 当 ℓ 过大时, 特征结构会消失.

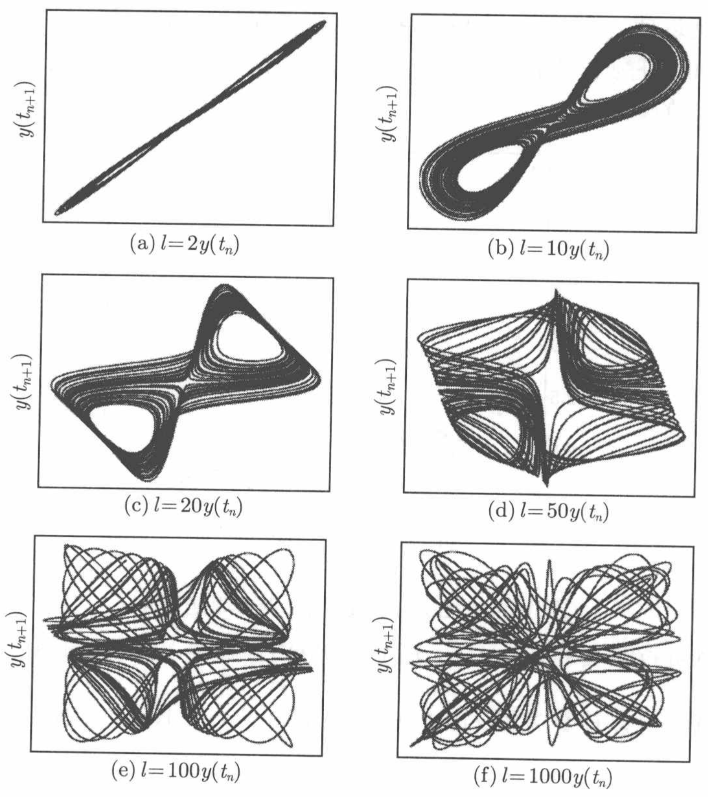


图 10.10 Lorenz 吸引子: 各种延迟长度 ℓ 的不同取值对应的标度时间序列 $y(t_n)$ 对 $y(t_{n+\ell})$ 的图

Fraser 和 Swinney (1986) 建议通过计算 ℓ 的一个函数——平均相互信息——来确定 ℓ 的取值, 其定义如下. 设 U 和 W 分别为取值为 $u_1, u_2, \cdots$ 和 $w_1, w_2, \cdots$ 的随机变量. 则平均相互信息函数定义为

$$H = \sum_i \sum_j \Pr\{U = u_i, W = w_j\} \log_2 \frac{\Pr\{U = u_i, W = w_j\}}{\Pr\{U = u_i\} \Pr\{W = w_j\}},$$

其中, $\Pr\{U = u_i, W = w_j\}$ 为 U 和 W 的联合概率. 对于一个连续标度时间序列 $Y(t_n)$, 可以将 $Y(t_n)$ 的取值范围分成一些小区间, 如 $I_1, I_2, \cdots$. 当 ℓ 取各种不同的值时, 我们对 $Y(t_n)$ 和 $Y(t_{n+\ell})$ 间的平均相互信息感兴趣. 从而, $H(\ell)$ 定义为

$$H(\ell) = \sum_i \sum_j \Pr\{Y(t_n) \in I_i, Y(t_{n+\ell}) \in I_j\} \log_2 \frac{\Pr\{Y(t_n) \in I_i, Y(t_{n+\ell}) \in I_j\}}{\Pr\{Y(t_n) \in I_i\} \Pr\{Y(t_{n+\ell}) \in I_j\}}.$$

对于可观测时间序列 $y(t_n)$, 定义区间 I_i, 并且让它的概率与访问区间 I_i 的次数成正比. Fraser 和 Swinney (1996) 建议对 $\ell = 1, 2, 3, \cdots$, 计算 $H(\ell)$, 并且将 ℓ 的值取为第一个最小值出现的地方, 用这个值作为方程 (10.1) 中定义的 $y^{(p)}(t_n)$ 的计算中的时间延迟值. 图 10.11 中就采用了这样的做法. 这里的计算中, 模拟了一个长度为 20,000 的序列, 并且 $y(t_n)$ 的取值范围被分成 200 个小区间. 第一个最小值出现在 $\ell = 16$ 的地方.

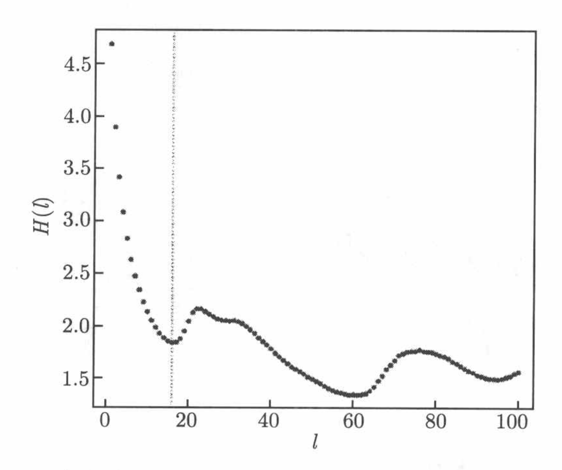


图 10.11　Lorenz 吸引子: 平均相互信息

Lorenz 吸引子的标度序列 $y(t_n) = x_1(t_n)$ 的平均相互信息: 第一个极小值出现在 $\ell = 16$ 处

取 $\ell = 16$, 可以计算方程 (10.1) 中定义的 $y^{(p)}(t_n)$. 图 10.12 画出了 $y^{(p)}(t_n)$ 的

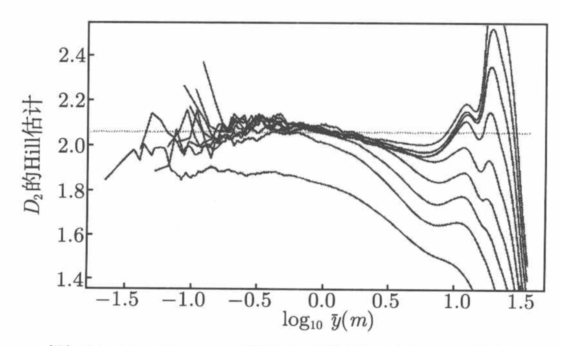


图 10.12　Lorenz 吸引子的嵌入的 D_2 估计

利用时间延迟的方法对 Lorenz 吸引子的 D_2 估计, 假设只有 $x_1(t_n)$ (如方程 (1.2)) 是可观测的. 这些曲线表示利用 $y^{(p)}(t_n)$ 所得到的维数估计, 其中 $p = 2, 3, \cdots, 10$. 水平点线的位置为 2.06

D_2 估计. 最底下的曲线表示 $y^{(2)}(t_n)$ 的维数估计, 往上一条曲线表示 $y^{(3)}(t_n)$, 以此类推, 最上面的曲线表示 $y^{(10)}(t_n)$. 随着 p 的增大, 维数估计似乎在与图 10.9 中 D_2 取同样值的地方变得饱和. 在得出对于某个给定的 p, 维数估计变得饱和这样的结论时需要十分谨慎. 随着 p 的增长, 需要在一个幂律行为显著的范围内, 用随之增长的取样间隔得到维数估计. 例如, 图 10.13, 它在白噪声的情况下, 画出了 p 从 1 到 5 的 D_2 的估计. 当 $p = 5$ 时, 维数估计曲线仅仅刚好到达 5.

10.5 一个过程是随机的, 还是决定性的?

简单的非线性方程组会产生极端复杂的行为, 并且其动力性质常常似乎是随机的. 对于很多观测时间序列, 导致这个系统的背后的本质过程是不清楚的. 例如, 惠灵顿的气候能部分的被关于日最高气温、日降雨量, 甚至仅仅日最高风速的时间序列刻画. 或者, 地球地壳构造板块的运动显然是由有记录的地震和地表形变引起. 然而, 在这两个过程中, 都仅仅只是观测到一个背后的复杂过程的某些明显的方面, 由此得到 §10.4.4 中关于投影的概念. 一个有趣的问题是, 给定一个关于某个背后的更本质的过程的时间序列, 能否确定这个背后的系统是由一组确定性的方程所支配, 还是一个随机的系统.

维数估计可以用来作为一种判断一个时间序列是随机的, 还是决定性的. 给定一个时间序列, 通常是单变量的, 对于 $p = 2, 3, \cdots$, 计算出方程 (10.1) 中的时间延迟嵌入. 通常对于每个 p, 都会估计关联维数 D_2. 如果时间序列是决定性的, 则 D_2 的估计将会趋于一个极限. 如果时间序列是随机的, 则 $D_2 = p$, 从而随着重建相空间的维数的增长, 维数估计会无限增长. 之所以如此是因为这样的假设: 随机时间序列具有无限多的自由度, 而由动力系统生成的时间序列只有有限多自由度.

在 Osborne 和 Provenzale (1989) 这篇重要的文章中, 与所期望的相反, 他们证明某个具有幂律谱有色的噪声过程在一个有限维饱和 (Theiler, 1991; Kennel, Isabelle, 1992). 其后, Cutler (1994) 证明, 他们的例子之所以不具有无限维, 是因为其幂律谱是离散的. 再后来, Cutler (1997) 更精确的定义了随机性和决定性的含义. 这会在这节中给出一个简要的综述.

10.5.1 定义 (Cutler, 1997)

1. 时间序列 $\{Y(t_n)\colon n = 0, 1, \cdots\}$ 被称为严格稳定的, 如果对任意有限族 $t_1, \cdots, t_n$ 及任意 τ,

$$\Pr\{Y(t_1) < y_1, \cdots, Y(t_n) < y_n\} = \Pr\{Y(t_1 + \tau) < y_1, \cdots, Y(t_n + \tau) < y_n\}.$$

2. 称映射 $g\colon \mathcal{X} \to \mathcal{Y}$ 满足 Lipschitz 条件, 若对任意的 $x_1, x_2 \in \mathcal{X}$,

$$\rho_2(g(x_1), g(x_2)) \leqslant k\rho_1(x_1, x_2),$$

其中 k 为常数, $\mathcal{X}$ 和 $\mathcal{Y}$ 为分别具有度量 ρ_1 和 ρ_2 的度量空间. 若 g 还是一一的, 并且 g^{-1} 也满足 Lipschitz 条件, 则称 g 为双 Lipschitz 的.

3. 设 $\{Y(t_n)\colon n = 0, 1, \cdots\}$ 为 $\mathcal{Y}$ 中取值的严格稳定的时间序列. 定义预言维数 ζ 为最小的整数 $n \geqslant 1$, 使得存在映射 $\Upsilon\colon \mathcal{Y}^n \to \mathcal{Y}$, 满足

$$Y(t_n) = \Upsilon[Y(t_0), \cdots, Y(t_{n-1})]$$

以概率 1 成立. 若对于任意 $n \geqslant 1$, 都不存在这样的函数 Υ, 则定义 $\zeta = \infty$.

由下面的定理可知, 一个具有已知的预言函数 Υ 和有限预言维数 ζ 的严格稳定过程, 可以被一个前 ζ 次观测的函数预言. 用这一点来定义随机性和决定性时间序列.

10.5.2 定理 (Cutler, 1997, 定理 2.2)

设 $\{Y(t_n)\colon n = 0, 1, \cdots\}$ 为一个具有有限预言维数 ζ 和预言函数 Υ 的严格稳定时间序列. 则对于任意整数 $m \geqslant 0$,

$$Y(t_{m+1+\zeta}) = \Upsilon\big(Y(t_{m+1}), Y(t_{m+2}), \cdots, Y(t_{m+\zeta})\big)$$

以概率 1 成立.

10.5.3 定义 (Cutler, 1997)

一个严格稳定时间序列 $\{Y(t_n)\colon n = 0, 1, \cdots\}$ 被称为决定性的, 若 $\zeta < \infty$; 被称为随机的, 若 $\zeta = \infty$, 其中 ζ 为预言维数.

由下面的定理知, 在某些条件下, 对于恰当的 p, 与具有不变测度 μ 的动力系统相关联的 Rényi 维数, 恰好等于与具有分布 P_n 的延迟时间序列 $Y^{(p)}(t_n)$ 相关联的维数.

10.5.4 定理 (Cutler, 1997, 推论 2.26)

设 $T\colon \mathcal{X} \to \mathcal{X}$ 为具有不变测度 μ 的动力系统, 即 $X(t_n) = T^n(X(t_0))$, 对于 $n = 1, 2, \cdots$, 其中 $X(t_0)$ 为具有分布 μ 的随机初始条件. 设投影 $\Pi\colon \mathcal{X} \to \mathcal{Y}$ Borel 可测, 并且 $Y(t_n) = \Pi[T^n[X(t_0)]]$ 为结果函数时间序列, 具有联合分布 P_n, 即

$$P_n(y_0, \cdots, y_{n-1}) = \Pr\{Y(t_0) < y_0, \cdots, Y(t_{n-1}) < y_{n-1}\}.$$

还假设对于某个 p, 由方程 (10.2) 定义的延迟坐标映射 Ψ_p 为双 Lipschitz 的, 并且 T 满足 Lipschitz 条件 (定义 10.5.1). 则

1. 函数时间序列 $Y(t_n)$ 是决定性的, 其预言维数 $\zeta \leqslant p$;

2. 对于任意 $n \geqslant p$, $D_q(\mu) = D_q(P_n)$, 对于 $p \in \mathbb{R}$, 其中 $D_q(\mu)$ 由方程 (2.6) 定义, 但需用 μ 替代分布 ν;

3. 对于任意 $n \geqslant p$, $\dim_{\mathrm{H}}(\operatorname{supp}\mu) = \dim_{\mathrm{H}}(\operatorname{supp} P_n)$.

10.5.5 例子——Gauss 时间序列

模拟一个标度 Gauss 时间序列 $\{y(t_n)\colon n = 0, \cdots, 100000\}$, 并且利用方程 (10.1) 计算 p-时间序列 $y^{(p)}(t_n)$ (注意到这里 $\ell = 1$). 图 10.13 显示了利用序列 $y^{(p)}(t_n)$ 对 D_2 的估计, 取 $p = 1, \cdots, 5$. 因为这个过程是随机的, $D_2 = p$. 然而, 注意到随着 p 的增长, 维数估计越来越多的受到边际效应的影响, 并且对于 $p = 5$, 实际上不再存在幂律关系, 因为估计范围刚刚达到接近 5 的程度.

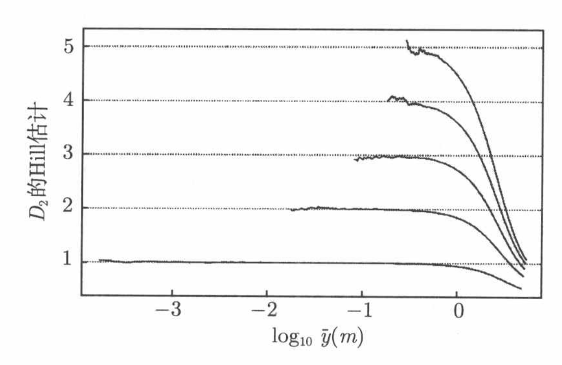


图 10.13 白噪声的嵌入的 D_2 估计

白噪声的标度时间序列 $y(t_n)$ 的 D_2 估计, $n = 0, \cdots, 100000$. 最下的曲线代表用序列 $y^{(1)}(t_n)$ 所得到的估计, 次下的表示 $y^{(2)}$ 对应的估计, 以此类推, 其中 $y^{(p)}$ 由方程 (10.1) 给出. 这里有 100 个自举抽样, 每个由 100000 个点间距取样组成

Wolff (1990) 在 $p = 2$ 时, 得到了关于一个自回归时间序列和一个移动平均时间序列的关联积分表达式. 他还利用 Gauss 白噪声新方法模拟了 AR(1) 和 MA(1) 过程, 并且计算了时间延迟嵌入 $y^{(p)}(t_n)$. 因为过程是随机的, D_2 应该等于 p, 即重构相空间的维数. 在 Wolff (1990) 的模拟中, 他所得到的估计为 $D_2 \leqslant p$, 两者的差值随着 p 的增大而增大; 即边际效应越来越不可忽视. 当模拟序列相邻两个观测间的自关联增大时, 对 D_2 的估计随之变小.

这些例子表明, 一个随机过程可能有无穷多个自由度, 从而具有无穷多维, 但边界效应使得人们无法从观测时间序列中察觉这一点. 随着 p 的增大, 边界效应变得越来越剧烈, 使得从表面上看来, 这一过程的维数趋向于一个有限的值, 正如对一个决定性时间序列所预期的那样.

10.6 具有幂律性质的随机过程

10.6.1 历史回顾

有三篇重要的文献提供了关于具有幂律比例关系和长程相关性的过程的开创性想法 (Hurst, 1951; Rosenblatt, 1961; Lamperti, 1962). 下面将简要介绍这些文献中包含的想法.

Harold Edwin Hurst (1951) 是一位水利学家, 他在 Aswan 研究尼罗河的排放率. 给定水库的流入量和流出量, 他导出了一个理想水库在给定时间范围 $(0, T)$ 内的库容范围表达式, 记为 $R(T)$, 确保水库可以安然度过丰水期 (即保证水库不超过库存容量), 并且满足枯水期的用水需求. 即 $R(T)$ 是水库在时间长度为 T 的范围内能正常发挥其功能的所需容量. 它依赖于时间长度, 因为可以认为所需容量与 Brown 运动的路径相似. 时间越长, 过程中涉及的变量越多, 从而各种极端情况出现的可能性就越大. Hurst (1951) 分析了历史记录, 通过对不同时期的历史记录的计算, 发现 $R(T)/S$ 近似等于 $(T/2)^H$, 其中 S 为 $R(1)$ 的标准差. Hurst (1951) 指出, 如果入库流量和出库流量是相互独立的, 类似于多个独立同分布随机变量的和, 则可以证明 R/S 满足一个这样的规律, 但 $H = 1/2$. 他发现 H 的典型取值大概是 0.73. Hurst 将这一结果归因于流量间具有连续的相关性. Hurst 参数 H 现在经常出现在各种关于具有长程相关性的随机过程的公式中. 更多的关于河流量随机化模型的信息可以参考 Lawrance 和 Kottegoda (1977) 及 Lloyd (1981). 所谓的 R/S 分析经常被用来估计 Hurst 参数, 关于这种方法的描述可以参见 Mandelbrot 和 Wallis (1969), 关于各种极限分布的结果可以参考 Mandelbrot (1975).

Rosenblatt (1961) 考虑了一列强混合的随机变量的和的分布. 称序列 Y_n 为强混合, 若

$$\left|\Pr\{Y_n \in A, Y_{n+m} \in B\} - \Pr\{Y_n \in A\}\Pr\{Y_{n+m} \in B\}\right| \leqslant d(m),$$

其中 $d(m)$ 趋于 0 当 $m \to \infty$. 他证明, 若这样的一个序列的和满足某些矩条件, 则它们的极限分布在适当的正规化后是一个 Gauss 分布. 他还给出了一个反例, 其中观测序列 Y_n 满足这些矩条件, 但是这个序列是严重相关的. 正规化后的和趋于一个非 Gauss 分布. 他推断序列 Y_n 不满足强混合条件. 事实上, 它显示了长程相关的一种形式.

称具有 0 期望的平稳序列 Y_n 落在过程 $X(t)$ 的吸引域内, 若当 $N \to \infty$,

$$X_N(t) = \frac{1}{d_N}\sum_{j=1}^{\lfloor Nt \rfloor} Y_j \tag{10.3}$$

的有限维分布趋于 $X(t)$ 的有限维分布, 其中 d_N 为任意正的正规化因子, 当 $N \to \infty$, d_N 趋于无穷, $\lfloor Nt \rfloor$ 表示不超过 Nt 的最大整数. Rosenblatt (1961) 的反例收敛到一个过程, 现在称其为 Rosenblatt 过程.

下面的讨论将用到有限维分布相等的概念. 称两个随机过程 $X_1(t)$ 和 $X_2(t)$ 具有相同的有限维分布, 若对任意 $n \geqslant 1$, 以及 $t_1, t_2, \cdots, t_n$,

$$\big(X_1(t_1), X_1(t_2), \cdots, X_1(t_n)\big) \stackrel{d}{=} \big(X_2(t_1), X_2(t_2), \cdots, X_2(t_n)\big),$$

其中 $\stackrel{d}{=}$ 表示概率分布相等. 将其简记为 $\{X_1(t)\} \stackrel{d}{=} \{X_2(t)\}$.

Lamperti (1962) 对过程 $X(t)$ 引入了比例的想法. 他定义一个 d 维过程 $X(t)$ 为半稳定的, 若它满足一个简单的连续性条件, 此外对于 $s > 0$, 关系

$$\{X(st)\} \stackrel{d}{=} \{b(s)X(t) + c(s)\}$$

成立, 其中 $b(s)$ 为一个正的函数, 并且 $c(s) \in \mathbb{R}^d$. 随后, Lamperti (1962) 证明, 如果 $X(t)$ 是一个适当的半稳定过程, 并且 $X(0) = 0$, 则 $c(s) = 0$ 且 $b(s) = s^H$, 其中 H 是一个正的常数. 即

$$\{X(st)\} \stackrel{d}{=} \{s^H X(t)\}, \tag{10.4}$$

这个式子现在被普遍地视为自相似随机过程的定义. Lamperti (1962) 还证明式 (10.3) 中的正规化常数 d_N 具有 $N^H L(N)$ 的形式, 其中 $L(\ )$ 为一个在无穷远处变化缓慢的正函数, 并且进一步, 极限过程 $X(t)$ 实际上是自相似的.

Taqqu (1975, 1977, 1978, 1979) 在一系列文章中考虑了一族极限自相似随机过程, 这些过程源于式 (10.3) 中给出的和式. 令 $Y_j = h_m(Z_j)$, 其中 $h_m(\ )$ 为 m 次 Hermite 多项式, Z_j 为具有长程相关性的随机变量组成的 Gauss 序列, 它们的相关性被参数化为一个 H 的幂函数. 则极限随机过程 $X(t)$ 是自相似的. 进一步, 当 $m = 1$ 时, 极限过程是分式 Brown 运动, 当 $m = 2$ 时, 是上面讨论过的 Rosenblatt 过程. Rosenblatt (1979, 1981), Dobrushin (1979), Dobrushin 和 Major (1979) 及 Major (1981) 还引入了类似于中心极限形式的定理. Taqqu 和 Wolpert (1983) 讨论了一个 Poisson 测度下的相似过程. 更多的细节可参考 Samorodnitsky 和 Taqqu (1994).

10.6.2 自相似随机过程

一个随机函数 $\{X(t)\colon -\infty < t < \infty\}$ 的增量被称为具有参数 H 的自相似, 若对任意 $s > 0$ 和任意 τ,

$$\{X(st+\tau) - X(\tau)\} \stackrel{d}{=} \{s^H(X(t+\tau) - X(\tau))\}.$$

若 $X(t)$ 的增量是自相似的, 并且 $X(0)=0$, 则 (10.4) 中的 $X(t)$ 也是自相似的. 若 $X(t)$ 具有自相似的、稳定的增量, 并且是均方连续的, 则可以证明 $0 \leqslant H < 1$.

协方差结构可以从上面的比例性质直接推导得到. 令 $X(t)$ 为具有平稳自相似增量的过程. 则协变函数为

$$\begin{aligned}&E[(X(t+\tau+1)-X(t+\tau))(X(t+1)-X(t))]\\&=\frac{1}{2}\sigma_H^2\{|\tau+1|^{2H}+|\tau-1|^{2H}-2|\tau|^{2H}\},\end{aligned}$$

其中, 对于所有的 t, $\sigma_H^2=E[(X(t+1)-X(t))^2]$.

过程 $X(t)$ 被称为各向同性的, 若

$$\{X(t)-X(s)\}\overset{d}{=}\{X(|t-s|)\}. \tag{10.5}$$

在这种情况下, $E[X^2(t)]=|t|^{2H}E[X^2(1)]$, 并且

$$E[X(t)X(s)]=\frac{1}{2}\{|s|^{2H}+|t|^{2H}-|t-s|^{2H}\}E[X^2(1)].$$

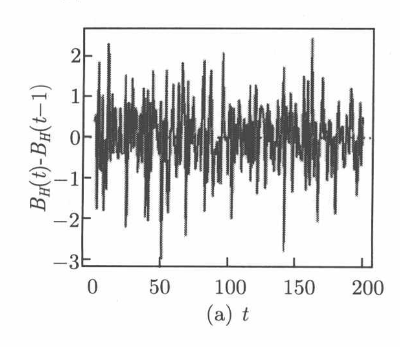
10.6.3 分式 Brown 运动

一个 Gauss 过程被它的协方差函数唯一确定. 这个唯一的 Gauss 自相似过程称为分式 Brown 运动, 记为 $B_H(t)$. 分式 Brown 运动的增量被称为分式 Gauss 噪声. 若 $B_H(0)=0$, 则过程 $B_H(t)$ 是各向同性的 (方程 (10.5)).

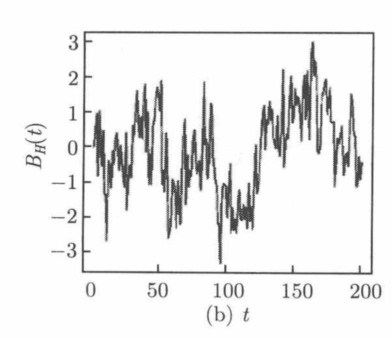
当 $H=\frac{1}{2}$, $B_H(t)$ 是 Brown 运动. 当 $H>\frac{1}{2}$, 自相关性为正, 并以指数速度下降, 从而长程相关. 当 $H<\frac{1}{2}$, 自相关性为负, 并具有很快的下降速度. 当 $\tau\to\infty$ 时,

$$E[(X(t+\tau+1)-X(t+\tau))(X(t+1)-X(t))]\approx H(2H-1)\sigma^2|\tau|^{2H-2}$$

(Mandelbrot 和 Van Ness, 1968). 这些特性可以在图 10.14 中清楚地看到, 其中增量过程中的正的自相关性对应于 $H=0.8$. 还注意到 $B_H(t)$ 的方差如何随着 H 的增大而增大, 与上面的各种关系一致.



(a) t



(b) t

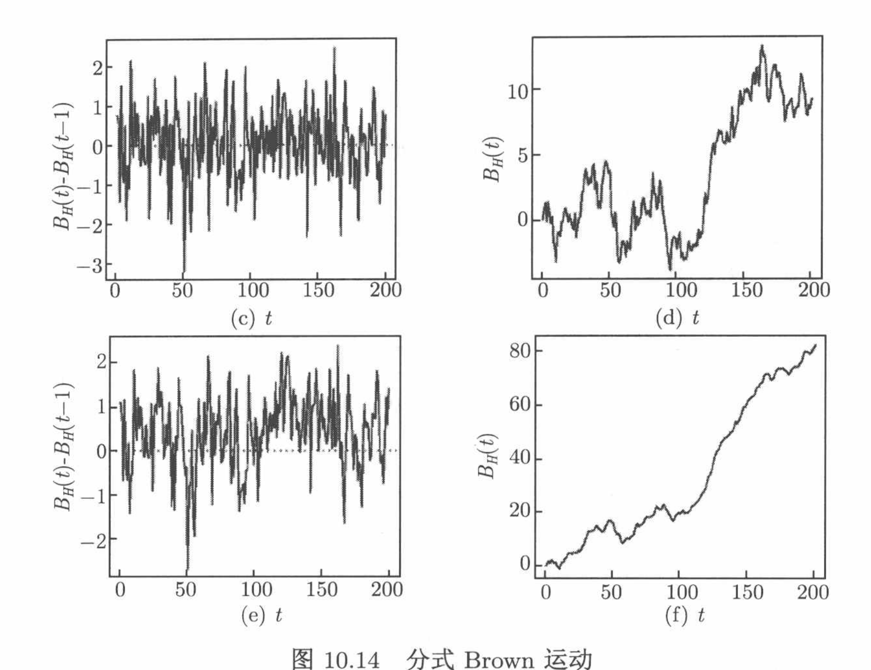


图 10.14 分式 Brown 运动

分式 Brown 运动的图, 其中 $H = 0.2, 0.5, 0.8$ ((b), (d), (f)), 对应的增量过程在左边. 这些序列是利用 Davies 和 Harte (1987) 中的方法模拟得到的

关于分式 Brown 运动的样本路径的性质有一系列的结果. 它是一个连续的过程但无处可微. Orey (1970) 证明其图像的 Hausdorff 维数为 $2-H$. 这里的图像即由点 $(t, B_H(t))$ 构成的集合, 其中 t 在一个有正测度的区间上取值. 从而, H 越接近 0, 图像越能填满平面 (图 10.14 中右图). 具有长程相关性的过程 $\left(\frac{1}{2} < H < 1\right)$, 其填充的空间较少.

图 10.15 画出了一个分式 Brown 运动的样本路径的二维图像. $\mathbb{R}^d$ 空间中的 Brown 运动的路径的 Hausdorff 维数 (及盒维数) 为 $\min(d, 2)$. 因此对于 $H = \frac{1}{2}$, 预期其图像可以填满局部区域. $d \geqslant 3$ 的情形见 Ciesielski 和 Taylor (1962), $d = 2$ 的情形见 Taylor (1964). 当 $H \neq 0.5$ 时, Mandelbrot (1977, §7, 284 页) 证明 $\mathbb{R}^d$ 中路径的 Hausdorff 维数为 $\min(d, 1/H)$. 这与 $H = 0.8$ 时的图像一致, 只能填充较少的空间. 其他附加过程的相关结果在 Berman (1970) 和 Marcus (1976) 中给出. 更多的讨论也可以参考 Adler (1981), Falconer (1990, 第 16 章) 和 Taylor (1986, 390 页).

Davies 和 Hall (1999), Chan 和 Wood (2000), Constantine 和 Hall (1994) 以及 Feuerverger 等 (1994) 利用上面的想法刻画了光滑表面的粗糙程度. 每相邻一个固定的间隔, 就测量一次表面的高度, 并假定测量是从一个类 Gauss 过程中取样,

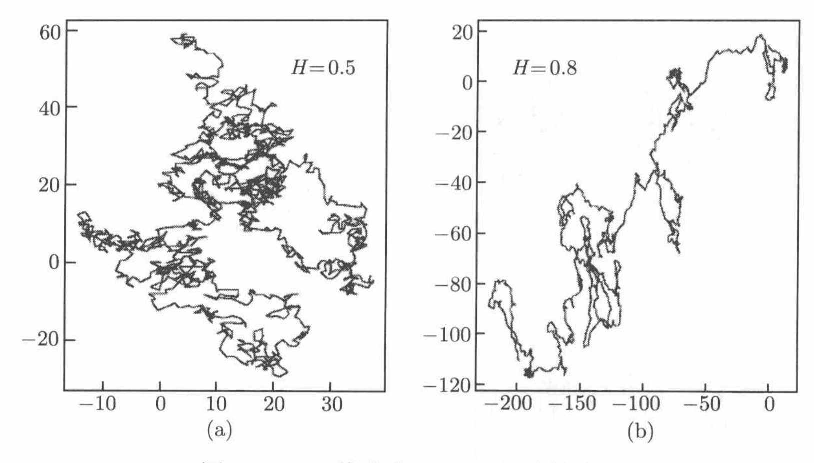


图 10.15　2 维分式 Brown 运动的路径

2 维分式 Brown 运动的路径, 其中 $H = 0.5$ 和 0.8, $t = 0, \cdots, 1500$. 这些序列是利用 Davies 和 Harte (1987) 中的方法模拟得到的

该过程的自协方差函数满足幂律. Orey (1970) 还确定了 Gauss 过程的水平交叉的 Hausdorff 维数, 并将它表示为一个自协变函数中的幂指数的函数. 从而通过估计自协变函数中的幂指数, 可以得到水平交叉的 Hausdorff 维数的估计. 这也被解释为表面粗糙程度的一个度量.

Davies 和 Harte (1987) 描述了为检验一个假设而作的检验统计, 这个假设认为一个取样时间序列是白噪声, 而不是分式 Gauss 噪声.

10.6.4　其他"幂律"过程

分数差异 ARIMA 过程 (Hosking, 1981) 也显示出长程相关性. 分数差异是通过在 ARIMA 模型中将后移位算子表示为一个无穷二项式级数得到的, 即

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k.$$

引入这样一个模型的部分动机是, 希望能提供一个灵活的框架, 能同时处理观测时间序列的短期和长期相关性. Li 和 McLeod (1986) 讨论了模型的适宜性及如何构造. Geweke 和 Porter-Hudak (1983) 比较了分数差异 ARIMA 模型和分式 Brown 运动, Haslett 和 Raftery (1989) 利用这一过程模拟了 Ireland 的风力资源, Porter-Hudak (1990) 还利用它模拟金融时间序列. Taqqu 等 (1995) 讨论了各种估计 Hurst 参数 H 和分数差异值 d 的方法. Beran (1994) 一书提供了关于长期记忆过程的有用的综述.

若一个过程满足 $X(0) = 0$, 其增量独立且满足稳定率, 则这个过程是自相似

的. 反之, 一个具有独立增量的自相似过程, 其增量必然满足稳定率 (Lamperti, 1962). 这些过程能被它们的无穷可分特征函数表示. 例如, 可参考 Kolmogorov 和 Gnedenko (1954), Lukacs (1960), Fristedt (1974) 以及 Breiman (1968). Fristedt (1974) 也包含了关于样本路径性质的结果.

10.6.5 "重分形"随机过程

最近, 一些幂律随机过程被描述为"重分形". 一个重分形随机过程被定义为一个与 $\theta(q)$, $\tilde{\theta}(q)$ 和 $\theta^{\star}(q)$ 类似的矩函数, 这三个函数分别见定义 2.4.1, 2.3.1 和 6.1.2.

令 $\mathcal{T}$ 和 $\mathcal{Q}$ 为实直线上具有正长度的区间, 使得 $0 \in \mathcal{T}$ 和 $[0,1] \subseteq \mathcal{Q}$. 一个随机过程 $X(t)$ 被称为重分形的, 若它具有稳定增量, 并且对于所有的 $t \in \mathcal{T}$ 和 $q \in \mathcal{Q}$, 有

$$E[|X(t)|^q] = c(q)t^{\theta^{\ddagger}(q)+1}, \tag{10.6}$$

其中, $c(q)$ 和 $\theta^{\ddagger}(q)$ 都是关于 $q \in \mathcal{Q}$ 的函数 (Mandelbrot et al., 1997).

现在考虑一个自相似随机过程, 它满足式 (10.4) 中给出的比例关系. 则

$$\{X(t)\} \stackrel{d}{=} \{t^H X(1)\},$$

从而

$$E[|X(t)|^q] = t^{qH} E\{|X(1)|^q\}.$$

若取 $c(q) = E\{|X(1)|^q\}$ 和 $\theta^{\ddagger}(q) = Hq - 1$, 则上式满足式 (10.6) 中给出的关系. 注意到 $\theta^{\ddagger}(q)$ 是 q 的线性函数. 根据前面的结果, 若重分形测度满足这个条件, 则所有的 Rényi 维数都一致, 从而, 即使这个测度所支撑的集合具有非常不规则的形状, 测度在这个集合上的分布在某种意义上是一致的. 某些时候, 称这样的测度为单分形测度. 在这种意义下, 一个自相似随机过程就是一个单分形过程.

一个满足多重比例定律的随机过程, 具有一个非线性的 $\theta^{\ddagger}(q)$ 函数. 假设过程 $X(t)$ 满足比例关系

$$\{X(ct)\} \stackrel{d}{=} \{W(c)X(t)\}, \tag{10.7}$$

对于所有的 t 和 $0 < c \leqslant 1$, 其中 $W(c)$ 是一个取正值的独立随机过程; 则

$$E\{|X(t)|^q\} = E[W(t)^q]E[|X(1)|^q].$$

进一步假设 $W(c)$ 满足下面的乘积关系:

$$\{W(ab)\} \stackrel{d}{=} \{W_1(a)W_2(b)\},$$

其中, W_1 和 W_2 是 W 的两个相互独立的拷贝, 并且 a 和 b 在区间 $(0,1]$ 中取值. 从乘积关系可得, 对于任意满足 $t_1t_2 = t$ 的 t_1 和 t_2,

$$E[W(t)^q] = E[W(t_1)^q]E[W(t_2)^q].$$

从而, $E[W(t)^q]$ 具有 $t^{\theta^{\ddagger}(q)+1}$ 的形式, 所以 $E[|X(t)|^q] = c(q)t^{\theta^{\ddagger}(q)+1}$, 其中 $\theta^{\ddagger}(q)$ 是一个关于 q 的非平凡的函数.

如何定义一个具有重分形性质的随机过程? Mandelbrot 等 (1997) 建议的方法是构造一个复合过程 $X(t) = B_H(T(t))$, 其中 $B_H(t)$ 是一个分式 Brown 运动, $T(t)$ 是一个关于 t 的连续的非减随机函数, 并且与 $B_H(t)$ 独立. 即有了一个在随机时间上的 Brown 运动. 特别地, 令 μ 为一个定义在区间 $[0,1]$ 上的重分形测度. 则一种办法是将 $T(t)$ 定义为 μ 的累积分布函数, 即

$$T(t) = \frac{\mu([0,t])}{\mu([0,1])}.$$

Mandelbrot (1999) 概述了一系列推动这个随机过程研究发展的历史模型. Heyde (1999) 也简要讨论了这个过程, 并指出这个过程的某些性质不适于对经济数据的模拟.

对于重分形测度的讨论, 我们强调某种形式上的整体平均和局部性质之间的联系. 在上面的讨论中, 我们用整体平均定义了重分形过程. 它的局部或样本路径性质通过式 (10.7) 中给出的比例关系来刻画. Riedi (2001) 讨论了所谓重分形谱和公式, 并把它和上面所讨论的矩函数联系起来.

Bouchaud 等 (2000) 分析了金融时间序列表面上的重分形行为, Fisher 等 (1997b) 利用这些模型描述德国马克和美元间的汇率. Mandelbrot 等 (1997) 和 Fisher 等 (1997a) 提出用这些过程作为财政收益的模型. Riedi 和 Willinger (2000) 建议用这些模型的变形来描述计算机网络通信.

第11章　地震分析

11.1 引　　言

这一章中,用地震产生的空间点模式来估计 Rényi 维数,并试着在一个合适的模型背景下对其进行解释. 这些分析过程是 §1.6 的继续.

地震是地壳的断裂过程,但可能不至于使地壳的表面裂开. 这些断裂过程可能在很深的地方出现,对地球的表面可能没有明显的影响. 地震经常因为板块移动而在临近板块的边界出现. 在某些情况下,当一个板块正在削弱另一个板块 (如在新西兰或日本的板块俯冲),或板块作水平滑动 (如在加利福尼亚的板块滑走) 时,地震就会发生. 也有很多板块内的地区经常发生地震,例如中国. 欧洲板块正被日本的太平洋板块的俯冲所挤压,压力被喜马拉雅山地区的印度板块所利用 (Sphilhaus, 1991).

地震是一种开始于时间和空间中某点的断裂过程. 然而,断裂的范围可能延伸到很远的距离,地震的等级与断裂的范围有关. Wells 和 Coppersmith (1994) 描绘了断裂的范围与地震的等级之间的不同关系. 地震的分类包括断裂开始的位置 (经度、纬度、深度)、起始时间、等级. 三维空间的定位称为震源,二维空间的定位称为震中.

估计地震产生的空间点模型的分形维数的动机就是,这个模型在某种意义上可能是自相似的. 就是说,点簇在更细的水平上是重复的. 也可以认为是,主要的断裂过程沿主要的断层出现,特别是板块的边界. 在主要的断层里面,有更小的断层出现,形成更小的断层网,于是可能生成某种自相似的网层,地震在某个时刻就集中爆发了. 然而,理论上有限个点的位置是零维的,于是,维数的估计究竟描述了什么特性呢?

维数估计并不是像样本均值那样对观测数据进行简单的统计,实际上,它描述了抽样数据的基本概率分布. 例如,例 10.3.1 中 Moran 瀑布过程的维数估计描述了 Moran 瀑布的概率测度,虽然维数估计只是基于有限个观测值. 在这点上有人必定会问:什么是地震的合适模型框架?在这个框架里怎样解释维数的估计?我们采用的是下面介绍的点过程框架.

11.1.1　点过程的远景

地震预测是一个非常复杂的问题,目前只处在初期阶段,它与天气预测有很多

的不同. 大部分的地震是观测不到的, 仅仅只有当地震的影响充分大时才能被观测到. 而且, 地震位置的确定有相当大的误差, 描述断裂过程的地球物理模型相当的初级.

不巧的是, 在地球物理的文献中, 分为两派：地球物理模型和统计模型. 我们采用的应该是这样一种模型, 它包含了已知的对地震的地球物理分析, 也能描述随机的影响, 它不能单一的被地球物理分支和统计分支所描述.

能够用来刻画地震的一个框架就是点过程模型 (Daley, Vere-Jones, 1988). 时间标度的 Poisson 点过程可以通过它的条件强度函数来刻画. 用 $N([t,t+\tau))$ 表示在时间段 $[t,t+\tau)$ 中地震发生的次数, 用 $\mathcal{H}_t$ 表示截止到时刻 t, 但不含 t 地震发生的历史. 条件强度函数定义为

$$\lambda(t|\mathcal{H}_t) = \lim_{\tau\to 0}\frac{1}{\tau}\Pr\{N([t,t+\tau)) > 0|\mathcal{H}_t\}. \tag{11.1}$$

于是, 在一个区间 A 中地震发生的期望数, 或强度测度为

$$E[N(A)] = \int_A \lambda(t|\mathcal{H}_t)dt.$$

地震过程自然用 t 来标度. 上面的定义能延拓到考虑空间的情形. 通常 $\lambda(t|\mathcal{H}_t)$ 是一种函数的形式, 它试着去描述所要求的地震的地球物理特征, 以及其他随机的特性. 它也包含一些需要估计的参数. 通过极大化对数似然估计, 强度测度可表示为

$$\sum_{i=1}^{N}\log\lambda(t_i|\mathcal{H}_t) - \int\lambda(t_i|\mathcal{H}_t)dt,$$

其中 N 表示观测到的地震发生的数目, t_i 表示每次地震发生震动的次数, $i = 1,2,\cdots,N$.

需要注明的是, 上面点过程的提法假定了地震在给定的时间和空间的某个点发生. 如果想包含那些断裂发生的范围和方向的相关信息, 可以加上一个标记. 关于加标记的点过程, 可以参考 Daley 和 Vere-Jones(1988). 正如已经注明的, 地震的等级与断裂的范围有关, 虽然这样的一个量的描述可能太简单, 不足以描述断裂的范围和释放的能量, 加上张量距等因素也许能得到一个更加有用的表示方法.

Vere-Jones(1999) 证明了, 在空间–时间点过程的情形, 对 Rényi 维数 (D_q) 的估计至少有两种不同的解释, 其中 $q\geqslant 2$ 是一个正整数. 首先, 空间区域固定, 在时间的增长周期上进行估计; 其次, 时间区间固定, 让空间区域的范围增长. 对这两种现象, Vere-Jones(1999) 对关联积分和关联维数的估计考虑了一致性的结果.

对空间区域固定的情形, 如果地震过程在时间上是平稳的和遍历的, 累积量满足某种正则条件, 则 Vere-Jones(1999) 证明了, Rénri 维数估计与地震过程的空间强度测度对应的 Rényi 维数是一致的.

对时间的观测周期固定的情形, 如果地震过程在空间上是平稳的和遍历的, 则 Vere-Jones(1999) 证明了, Rénri 维数估计与 Palm 分布的距测度的初始幂率增长一致.

在这一章中, 估计了地震空间位置的 Rényi 维数, 并从点过程模型的远景对维数进行了讨论. 这一章的分析基于 Harte(1998) 早期的工作.

11.2 数据来源

11.2.1 地震位置的确定

当地震发生的时候, 突然释放的巨大能量产生了一个 P 波 (主要的) 和一个 S 波 (次要的), P 波是纵向的, S 波是横向的, P 波比 S 波以更快的速度传递. 这些能被地震站检测到, 从每一个站点的地震仪描出的点的痕迹中, 可以确定 P 波和 S 波到达的时间、振幅, 以及持续时间. 用一个"速度模型"来描绘 P 波和 S 波在地震附近的地质结构的类型中传递的速度, 从三个或更多记录这次地震的地震站中的描出的点的痕迹, 就可以确定地震的位置 (断裂的初始点) 和地震发生的时间. 位置算法采用迭代最小二乘重加权程序, 将原点调整, 尽量减少剩余量. 一些地震分类中包含了位置估计程序里面每一维的标准差.

涉及到地震研究的空间物理学家也用传递时间和地震波的其他特性来确定地球的地质结构. 这些知识也能用来得到和修改速度模型. 确定地震的某些信息, 特别是深度和等级, 是高度相关的. 一个高估深度的速度模型往往低估等级, 反之也如此. 经常地, 从拟合程序中得到的剩余量包含一个系统的误差分量. 对于发生在离地震站有一定距离的地震, 如海洋中的, 系统误差可能会很大 (见 Harte, Vere-Jones, 1999). 于是, 在估计地震位置时, 标准差对绝对误差可能并没有给出一个精确的量度. 用 x 表示地震发生在 x_0 处的震源的估计位置, 则

$$x = x_0 + S(x_0) + \varepsilon,$$

其中 $S(x_0)$ 是系统误差分量, ε 是均值为零的噪声. 标准差将不会描述系统分量的范围, 但会简单地描述 $\mathrm{Var}(\varepsilon)$.

在我们的分析中, 将用两种地震分类, 关东分类和惠灵顿分类 (在 §1.6.1 中已经介绍过).

11.2.2 关东地震分类

关东地震分类包含日本的关东地区 (东京及其周边区域) 的地震. 这种分类由国家地球科学和灾难预防研究院的 Tsukuba-Shi 所提出. 对这种分析, 选择的是发生在东经 138.6° 到东经 141.9°, 北纬 34.6° 到北纬 37.3°, 等级 $\geqslant 2$, 1979 年 7 月 1 日到 1999 年 6 月 30 日之间的地震.

关东地区位于三个地震板块边界的交叉处, 就像一个“T”型交叉 (Sphilhaus, 1991). 西部与欧洲板块相邻, 这个板块的边界将日本向台湾方向拉长. 在关东区域里面, 它又被太平洋板块向东推移, 西南方向是菲律宾板块, 它同样受太平洋板块所推移. 这些边界在图 11.1 和图 11.2 中能部分显示出来. 两个图中的下左方是地震的一个边界, 从北–西北方向向南–东南方向运动, 这是太平洋板块与菲律宾板块之间的边界. 在图的顶部的左边的其他边界, 从西–西南方向向东–东北方向运动, 标出了太平洋板块与欧洲板块之间的边界. 关东地区的情况比惠灵顿地区的情况更复杂, 密集的地震有向更深的地方发生的趋势. 然而, 可以看到, 大部分深度大于等于 80 公里的地震与俯冲过程有关, 于是那些更浅的地震好像有更广泛的分布. 对于关东分类, 用 80 公里作为中等深度地震与深地震的一个分界.

图 11.1～图 11.3 分别是深地震、中等深度地震, 以及浅地震的震中所描的点, 图 11.4 表示太平洋板块推移菲律宾板块. 这个推移区域的地震描绘出了推移板块的摩擦边界的轮廓, 比惠灵顿区域 (图 1.6) 的情形更加清晰. 这可能预示着在关东分类中地震出现更少的位置误差. 同时需要说明的是, 像惠灵顿分类一样, 浅地震越来越聚集在一起. 与惠灵顿地区不同的是, 关东地区包含了活跃的火山区, 许多地震可能跟火山有关, 特别是伊豆半岛上的地震.

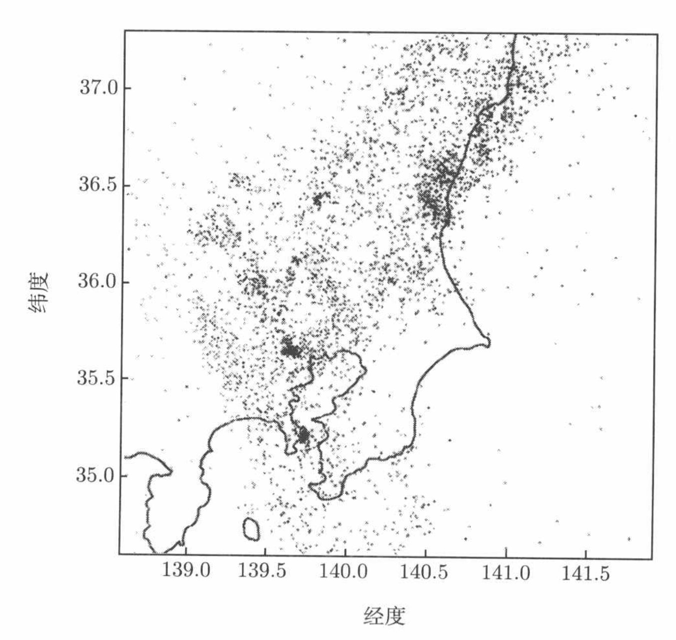


图 11.1　关东地震震中：深地震 (经度和纬度)

1985～1994 年的关东地震震中, 等级 ⩾ 2, 深度 ⩾ 80 公里. 最深的地震位于灰色阴影中最亮的部分, 最浅的地震位于最黑的部分. 描的点包含了 7570 个地震

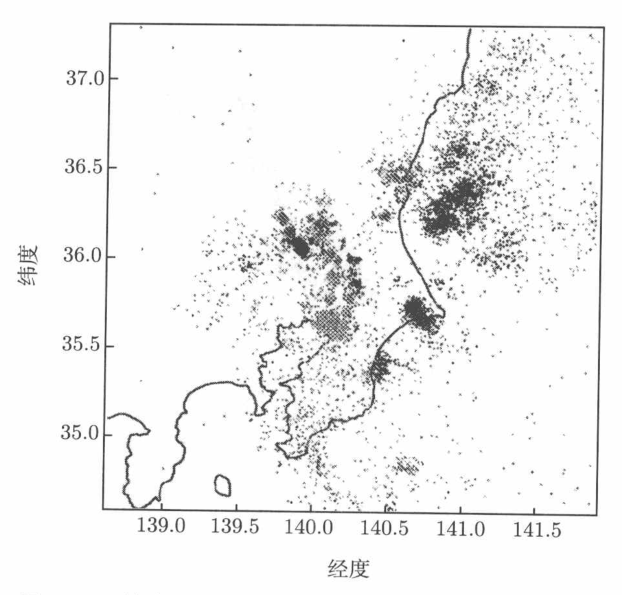


图 11.2 关东地震震中：中等深度地震 (经度和纬度)

1985～1994 年的关东地震震中, 等级 ⩾ 2, 40 公里 ⩽ 深度 < 80 公里. 最深的地震位于灰色阴影中最亮的部分, 最浅的地震位于最黑的部分. 描的点包含了 12486 个地震

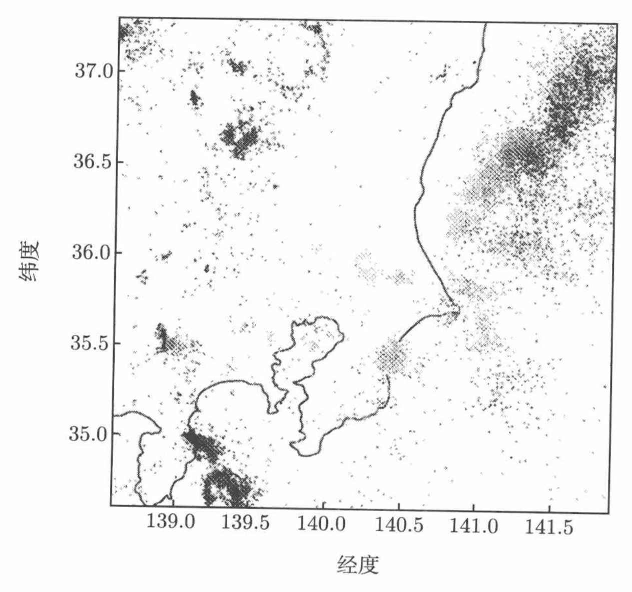


图 11.3 关东地震震中：浅地震 (经度和纬度)

1985～1994 年的关东地震震中, 等级 ⩾ 2, 深度 < 40 公里. 最深的地震位于灰色阴影中最亮的部分, 最浅的地震位于最黑的部分. 描的点包含了 23195 个地震

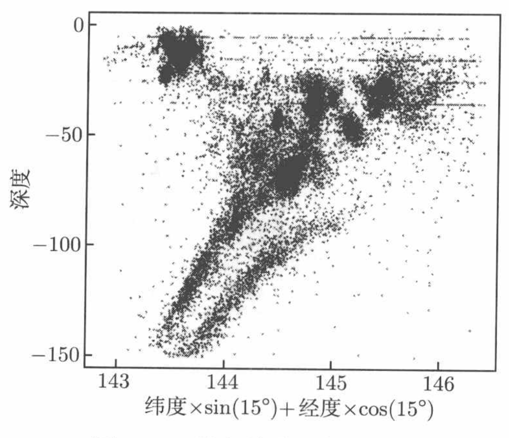


图 11.4　关东地震深度的截面

1985～1994 年的关东地震位置的深度的截面, 等级 $\geqslant 2$, 深度 < 150 公里, 北纬 36.1° 以南. 描的点包含了 23273 个地震. 画面展示了太平洋板块挤压菲律宾板块. 观测的视角大概是从南–东南方向向北–西北方向

11.3　引起偏差的影响

11.3.1　位置误差

最严重的一个问题就是地震的位置误差 (即数据中的噪声). 当点间距趋向于零时, 这会导致震中的维数估计趋向 2, 震源的维数估计趋向 3, 就像图 9.3 中描绘的同样的影响. 不太清楚怎样去纠正它, 就像加进去的噪声能有效地破坏数据中包含的任何无穷小的信息那样. 在目前的分析中, 一个复杂的因素是, 这个误差在任何被分析的地区都不是固定的. 在分类中, 制表式的标准误差对震源的精度是一个误导, 它们实际上描述的是给定地震的速度模型的拟合的精度, 不可能反映系统的偏差.

Papanastassiou 和 Matsumura (1987) 研究了关东分类中的震源位置的误差. 在那个时候, 地震网由 67 个站点构成：17 个地面站, 2 个隧道站, 45 个浅的炮眼站, 以及 3 个深的炮眼站. 他们有兴趣去确定第四深的炮眼站的最优位置. Morandi 和 Matsumura (1991) 改进为将检测能力看做是等级的一个函数. 在所分析的地区的东北角, 位置的误差会增大, 这不仅是因为这个区域 (大海) 站点很少, 而且跟这个区域不同的地质结构有关. 西边的站点比东边的站点更加敏感, 因为西边的区域是山区, 坚硬的岩床产生很低的地表噪声.

Papanastassiou 和 Matsumura (1987) 对地震的纬度、经度、深度以及发生的时间提供了位置误差的标准差的等高线图. 他们建议在所分析的区域中大部分地区

纬度的误差的标准差低于 1.5 公里, 东部 (沿海岸线) 增加到 3 公里. 经度的标准差更大一点, 虽然大部分地区不超过 2 公里, 但东部增加到 3 公里以上. 最大的误差出现在深度变量中, 深度越深, 误差越大, 东部更增加约 2 到 7 公里.

在惠灵顿分类中, 地震位置的精度也有类似的变化, 沿海岸线的地震和深地震精度会更差.

11.3.2 边界和缺项

我们的方法是在随机选择的地震中计算第 q 阶点间距. 所研究的地区强加边界的 ε 距离内发生的地震很少形成比 ε 更大的差异, 于是对大的 m 的值, $\hat{\theta}_m$ 会太小 (式 (8.8)). 这称为是边界的影响. 从式 (9.1) 给出的限制区域 $\mathcal{A}_\varepsilon$ 的每一对中选择一个点, 则能减少偏差. 更多的细节已经在 §9.2 中给出. 这种纠正在目前的分析中并没有被实施. 分析时所用的边界采用的是长方形的区域, 但所分析的区域中实际的地球物理边界是更加不规则的, 并没有提供这样的一个纠正.

当 q 阶点间距趋向于零时, Rényi 维数被定义成一种极限. 实际上, 维数是通过一段充分平的曲线估计出来, 需要假定当 y 充分小时, $\theta(q)$ 存在, 并且关联积分 $F_Y(y)$ 有一个指数为 $\theta(q)$ 的幂率变化, 即存在 ε, 对所有的 $y \leqslant \varepsilon$, $F_Y(y)$ 接近以指数为 $\theta(q)$ 的幂率变化. 然而, 如果边界和位置误差的影响比较严重, 则他们的影响会合并, 使得对所有 y 的值, $F_Y(y)$ 都会递减, 看不到基本的幂率变化.

边界影响的另一种类型由地震簇引起. 当 y 的值比簇的直径大, 但小于簇间距时, 就会引起抽样的点差减少, 从而引起 $\hat{\theta}_m$ 递减. 就像图 6.1 中的那种情形, 簇在簇的里面重复出现. 如果自相似的比例相同, 则这种自相似的行为在维数估计中表现出一种缺项周期; 如果自相似的比例不同, 则缺项周期有彼此不同的相位, 就会彼此抵消影响, 使得维数估计的曲线变得很平坦 (见图 10.5).

11.3.3 数据的四舍五入和变换

对数据进行四舍五入, 点间距被限制到相当少的一些离散的点时, 类似于缺项的影响也会出现. 将这看做是四舍五入的影响, 已经在 §9.3 中介绍过了. 在这种情况下, 式 (8.8) 给出的 $\hat{\theta}_m$ 也会有一种震荡行为, 与缺项的情形不同的是, 震荡的周期在未转化的 y 尺度上是恒定的. 四舍五入引起的另一个问题是, 不同的地震过程由相同的空间坐标来表示, 这样会出现零距离.

有时有两种不同的四舍五入方式. 例如, 在两种分类中, 大部分的地震的深度都是四舍五入到百米精度. 然而, 当地震记录站所记录的地震中, 浅地震出现很多时, 地震的深度很难区分. 这时, 惠灵顿分类中地震的深度限制在 5, 12, 或者 33km; 关东分类中地震的深度限制在 5, 15, 25 以及 35 km.

两种分类都将震源列表在球极坐标中, 即经度和纬度分别作为格林威治以东, 赤道以南 (或以北) 的度数, 以 km 作为单位的深度表示地表以下. 经度和纬度也都转成以 km 为单位. 于是所有用来计算的点间距都以 km 为单位, 这可能使得在

所分析的区域的端点处略有失真.

11.4 结　　果

对于惠灵顿分类和关东分类, 估计了不同深度的地震的 Rényi 维数, 分别计算了 $D_2, \cdots, D_5$. 在惠灵顿分类中, 地震深度分成了两级, 浅的和深的, 分别对应深度小于 40 公里和深度大于等于 40 公里. 关东分类中, 地震深度分成了三级, 浅的、中等深度的和深的, 分别对应深度小于 40 公里, 大于等于 40 公里小于 80 公里, 以及大于等于 80 公里. 对于惠灵顿分类法, 浅的和深的地震的维数估计在图 11.5 和图 11.6 中分别用点描出了. 而对于关东分类法, 三种深度的地震的维数估计分别在图 11.7~图 11.9 中用点描出.

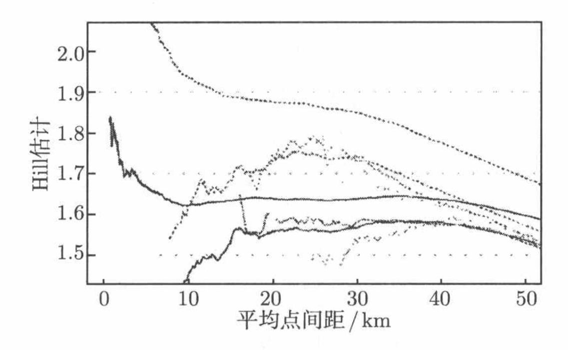


图 11.5　惠灵顿维数估计: 浅地震 (平均点间距和 Hill 估计)

浅地震 (深度小于 40 公里) 的 Rényi 维数的估计, D_2(黑色), D_3(黑灰色), D_4(浅灰色), D_5(最浅的灰色). 实线表示震中的估计, 虚线表示震源的估计 ($k = 100, n = 20,000, N = 15,228$)

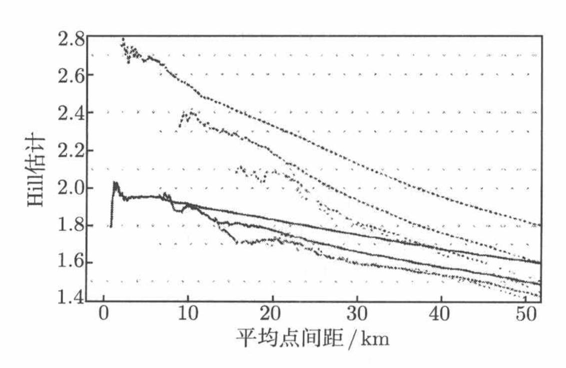


图 11.6　惠灵顿维数估计: 深地震 (平均点间距和 Hill 估计)

深地震 (深度大于等于 40 公里) 的 Rényi 维数的估计, D_2(黑色), D_3(黑灰色), D_4(浅灰色), D_5(最浅的灰色). 实线表示震中的估计, 虚线表示震源的估计 ($k = 100, n = 20,000, N = 6,746$)

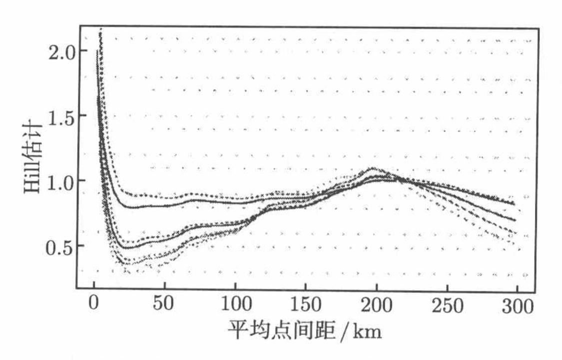


图 11.7 关东维数估计: 浅地震 (平均点间距和 Hill 估计)

浅地震 (深度小于 40 公里) 的 Rényi 维数的估计, D_2(黑色), D_3(黑灰色), D_4(浅灰色), D_5(最浅的灰色). 实线表示震中的估计, 虚线表示震源的估计 ($k = 100, n = 40,000, N = 38,566$)

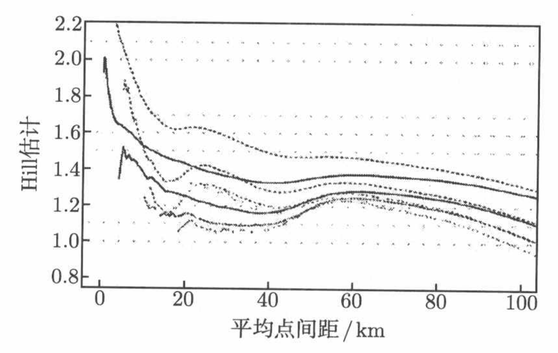


图 11.8 关东维数估计: 中等深度地震 (平均点间距和 Hill 估计)

中等深度地震 (深度大于等于 40 公里小于 80 公里) 的 Rényi 维数的估计, D_2(黑色), D_3(黑灰色), D_4(浅灰色), D_5(最浅的灰色). 实线表示震中的估计, 虚线表示震源的估计 ($k = 100, n = 25,000, N = 24,097$)

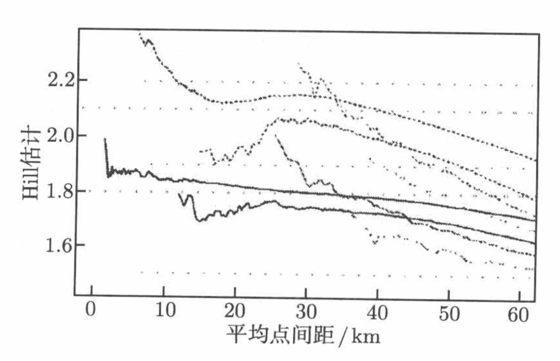


图 11.9 关东维数估计: 深地震 (平均点间距和 Hill 估计)

深地震 (深度大于等于 80 公里) 的 Rényi 维数的估计, D_2(黑色), D_3(黑灰色), D_4(浅灰色), D_5(最浅的灰色). 实线表示震中的估计, 虚线表示震源的估计 ($k = 100, n = 15,000, N = 13,620$)

在估计维数时, 希望在描点时找到一个相对平坦的区间, 使得在这个区间里面, 点间距的概率函数以某种幂率变化. 不幸的是, 当点间距轻微递减时, 噪声的影响使得所描的曲线递增到相空间的维数; 同时, 当点间距增大时, 边界的影响又使得曲线快速递减. 当这两种偏差都比较严重时, 它们就混在一起, 找不到明显的平坦的区间. 将维数的估计列在表中, 选择的区间是使得变化的幂率是我们估计的或期望的那样, 采用的是区间上的平均估计, 曲线的斜率也列出来了. 斜率的绝对值越大, 表示维数的估计在噪声和边界的双重影响下越混淆, 于是也就不确定. 斜率也就给出了维数估计受噪声和边界影响的一个粗略的刻画. 当某些点处曲线的斜率比较平或者有明显的斜坡, 它就刻画了一个图的尺度, 于是斜率对于比较不同的点也就给出了一个稍微客观的依据.

11.4.1　惠灵顿分类

浅地震的 Rényi 维数的估计在图 11.5 中用点描出. 数据集包含那些深度限制在 0, 5, 12 以及 33 公里的地震. 在给出的 15228 个浅地震中, 有 942 个地震的深度限制在这些值. 包含或排除这些地震都不会有明显的差别. 用 5~50 公里之间的点间距, 震中的 D_2 估计位于 1.60~1.68, 大约 1.64. 对于 D_3 和 D_4, 在 15~45 公里出现了一个平坦的区域, 给出的估计大约 1.57. D_5 的估计是不稳定的, 但好像与 D_3 和 D_4 的类似. 震源的估计更加不稳定, 这并不奇怪, 因为区域的深度 (40 公里) 相对于宽度和广度来说是非常浅的, 于是边界的影响就会强加一个更大的问题. 而且, 深度对于空间坐标来说, 可能是最不好确定的, 因此噪声就会影响到描点, 使得比震中的情形出现更大的点间距. 正因为这些, 我们希望找到一个更小的区间来显示幂率行为. 所做的估计和尺度区间在表 11.1 中都列出来了.

深地震的 Rényi 维数的估计在图 11.6 中用点描出, 一个平坦的区域并不明显. 深地震主要与俯冲板有关, 俯冲板从西南方向到东北方向穿过整个区域, 就像图 1.6 所展示的那样. 当震源被投影到地球的表面时, 震中就是地震的位置. 最终活动的震中面积相对比较小, 于是边界的影响相对于小的点间距来说就成了一个问题. 甚至震源的位置也被包含在相对小的区域中, 维数本质上就是俯冲板的宽度, 这就使得深地震可能有更大的位置误差. 对这两个问题, 当它们混在一起影响时, 并不奇怪, 这会使得维数估计的曲线在整个点间距的取值中呈现下降的趋势. 如果这两种偏差的影响不是很严重, 我们希望在 5~35 公里有幂率行为出现. 在这个区间中, D_2 的估计在 1.95~1.68 变化, 中点大概是 1.8. D_3 至 D_5 的估计线大约是平行的, D_3 比 D_2 接近低了 0.1, D_4 和 D_5 大约低了 0.15, 见表 11.1.

当点间距增加时, 表示深的震源的估计甚至递减得更快. 实际上没有平坦的区域, 虽然这部分地因为点的纵向变化. 对 D_q 的每一个估计, 在 10~30 公里的点间距的区间中取估计的中点. 这些值在表 11.1 中已经列出了.

表 11.1 基于惠灵顿分类的 Rényi 维数的汇总

惠灵顿分类			尺度值 (公里)	估计	斜率的绝对值
浅地震	震中	D_2	5~50	1.64	≈ 0
		D_3	15~45	1.57	≈ 0
		D_4	15~45	1.57	≈ 0
		D_5	35~45	1.57	不稳定
	震源	D_2	15~30	1.87	≈ 0
		D_3	20~30	1.75	≈ 0
		D_4	20~30	1.75	≈ 0
		D_5		1.70	不稳定
深地震	震中	D_2	5~45	1.80	0.008
		D_3	5~45	1.70	0.008
		D_4	20~40	1.65	0.008
		D_5	20~40	1.65	0.008
	震源	D_2	5~50	2.25	0.020
		D_3	10~50	2.00	0.020
		D_4	10~50	1.80	0.020
		D_5	35~45	1.80	0.020

注: 斜率是曲线在尺度值上的梯度, 斜率越大表示噪声和边界影响产生更大的混淆.

11.4.2 关东分类

浅地震 (深度小于 40 公里) 的 Rényi 维数的估计在图 11.7 中已经描出了, 只有 D_2 的估计表现出了幂率行为. 所有的估计起初都快速递减, 然后在大约 200 公里处递增到一个峰值. 对于这种点间距, 在震中和震源的估计中没有本质的区别. 这是因为 200 公里比深度维数的整个宽度大得多. 甚至对于小的点间距, 震中和震源的维数估计的差别看起来也是很小的, 这主要是因为图的比例尺的问题, 而且这与图 11.5 中惠灵顿浅地震也大致是可比的. 出现在大约 200 公里处的维数估计的峰值可能是由两个或更多的活跃的地震簇之间的距离所引起, 这些地震簇有效地控制了整个描点的特性. 图 11.7 中 D_2 的维数估计也支持了与图 10.6 中 β-分布的维数估计的一些相似性, 这与相对少控制的地震簇是一致的.

中等深度地震 (深度大于等于 40 公里小于 80 公里) 的 Rényi 维数的估计在图 11.8 中已经描出. 注意到表示震中的线在大约 60 公里处有一个局部极大值, 而表示震源的线在大约 25 公里和 60 公里处有两个局部的极大值, 这可能是由大量的地震簇所引起.

深地震 (深度大于等于 80 公里) 的 Rényi 维数的估计在图 11.9 中已经描出. 就像惠灵顿地区的深地震一样, 没有区间使得在该区间上的曲线是平坦的. D_2 的震源的估计显示了大约在小于 20 公里处有一个局部极小值, 大约 30 公里处有一个局部的极大值. 这可能代表了两个俯冲板块的摩擦边界的距离, 局部的极大值由

相反边界中选出的点对所引起. 维数的估计在表 11.2 中已经汇总, 分别列出了所选择的尺度区间、维数估计的平均值, 以及曲线的斜率, 就像在惠灵顿分类中所做的一样.

表 11.2 基于关东分类的 Rényi 维数的汇总

关东分类			尺度值 (公里)	估计	斜率的绝对值
浅地震	震中	D_2	25~150	0.80	0.001
		$D_3 - D_5$			不确定
	震源	D_2	25~150	0.90	≈ 0
		$D_3 - D_5$			不确定
中等深度地震	震中	D_2	15~80	1.40	下降
		D_3	15~80	1.35	下降
		D_4	15~90	1.15	下降
		D_5	20~90	1.15	下降
	震源	D_2	15~45	1.60	不确定
		D_3	15~45	1.55	不确定
		D_4	15~45	1.35	不确定
		D_5	20~45	1.35	不稳定
深地震	震中	D_2	5~55	1.80	0.002
		D_3	15~45	1.75	0.002
		D_4	30~55	1.75	0.008
		D_5	40~60	1.60	0.008
	震源	D_2	15~35	2.15	下降
		D_3	15~40	2.00	不稳定
		D_4	30~60	2.00	0.015
		D_5	40~60	1.80	0.015

注: 斜率是曲线在尺度值上的梯度, 斜率越大表示噪声和边界影响产生更大的混淆.

11.5 结果的比较和结论

两个地区的深地震的维数估计十分相似, 以同样的方式受边界和噪声的影响. 这由表 11.1 和表 11.2 中相似的斜率暗示出来. 同时关东地区包含了活跃的火山区, 而惠灵顿地区则没有. 然而, 惠灵顿地区和关东地区的深地震分别表示的是深度大于等于 40 公里和深度大于等于 80 公里的地震. 火山在更浅的深度会诱导地震的发生, 而所分析的深地震一般又与两个地区的板块俯冲有关.

两个地区的浅地震的维数估计是非常不同的. 在惠灵顿地区, 维数估计在很长的一段距离中显示出了幂率的变化, 而在关东地区, 这种变化仅仅对 D_2 是显然的, 而且只有当点间距大于 25 公里时才出现. 由于浅地震的最大深度是 40 公里, 我们希望边界的影响在 25 公里处就开始能被注意到. 这也是当点间距大于 25 公里时,

震中和震源的维数估计差不多相同的原因. 而且, 基于关东分类的维数估计都小于基于惠灵顿分类的维数估计, 这可能暗示了关东地区的地震的空间模型比惠灵顿地区的模型由更大密度的地震簇所刻画. 这也可能是惠灵顿地区没有火山, 而关东地区包含了活火山. 火山活动生成的地震簇会更小更密, 从而不能对关东地区的浅地震的 $D_3 - D_5$ 给出清晰的估计.

基于关东分类的中等深度地震的维数估计并不是单调递减 (当点间距递增时) 到其他描点所展示的同样的程度, 它们在大约 25 公里和 60 公里处包含两个局部的极大值. 关东地区的中等深度地震的震源的 $D_3 - D_5$ 也不能得到清晰的估计.

当震源的深度递增时, 两个地区所估计的维数都递增, 这与图 1.8 和图 11.1 中更多的一致空间模型是一致的. 这可能部分地因为位置误差的结果. 这些都强烈地依赖从震源位置看地震仪网络中有效站点的视角. 因此, 将期待深地震比浅地震有更大的位置误差. 这种假设也与深地震相对于浅地震而言, 维数估计描点过程中产生的更大的斜率是一致的.

11.5.1 点过程的设置

缺项由测度的支撑集所引起, 这个支撑集是一个有相同自相似比例的自相似集. 这使得当点间距以对数的尺度来描点时, 维数的估计出现一个周期性. 当自相似的比例不同时, 所诱导的缺项有不同的周期不同的相位, 相互抵消, 维数的估计就产生了一个平滑的曲线. 支撑在一个不同比例的自相似集上的空间强度测度, 对惠灵顿地区的浅地震将会产生一个相似的维数估计. 这样一个自相似的假设好像也是合理的, 这与应力分布的自相似特性有关. 采用不同的自相似比例也是与该地区的地质非均匀性是一致的.

然而, 如果有一种地震分类能包含几千年来的所有地震, 我们希望能得到一个什么样的维数估计值? 从分析中, 好像给出的空间中的地震簇在时间上也是聚集在一起的. 于是, 看一个给定时间周期的震中点, 就会看到一个地震簇; 在另一个不重叠的时间区间中, 又会看到在第一个区间中不存在的地震簇, 反之也是如此. 似乎在一个给定时间区间中剧烈活动的热点最终会消失, 新的热点又在其他地方出现. 在一个火山正在活动的火山区中, 这种现象会出现, 然后又消失, 然后新的火山又出现, 虽然从时间上来看会比较长. 可以想象, 如果这些热点随机出现, 一致活动, 直到应力在那个地区充分释放, 或者转移到临近的区域 (Lu et al., 1999; Bebbington, Harte, 2001), 有一种说法是, 在很长的一段时间中, 断裂的过程将充满整个空间, 于是, 震中和震源的 Rényi 维数会分别趋向于 2 和 3. 如果这个说法是对的, 那么在点过程的背景下, 怎么解释维数的估计?

这种说法与第二种情形更接近一些, 就是 Vere-Jones(1999) 所讨论的那种情形, 他们让空间区域固定, 让时间无限期增长, 此时 Rényi 维数的估计与地震过程的空

间强度测度对应的维数是一致的. 而在第一种情形中, 时间周期是恒定的, 空间区域增长. 这似乎是可信的, 因为区域的大小本质上是不变的, 由地球物理过程本身所决定.

我们提出的点过程通过一个条件强度函数来刻画, 式 (11.1) 已经给出这个函数, 这个函数与地震的历史过程有关. 考虑一个带条件强度的模型, 这个强度更多地依赖刚刚发生的地震, 风险最终变成与很久以前发生的地震无关. 因此, 现在的应力分布以及空间强度就能很好地支撑在一个自相似集上, 因为它们不依赖整个地震的历史. 空间强度测度可以随机地随时间演变, 虽然保留了自相似的结构, 但是由于远古的地震与当前的活动无关, 于是不会影响到当前的空间应力分布. 当估计 Rényi 维数时, 有一种可能就是, 根据给定模型指定的当前的影响, 对抽样的点间距进行加权.

如果有一种分类能包含几千年的数据, 我们就能及时地估计不同时间点的空间强度测度对应的 Rényi 维数, 对于所分析的时间点, 根据它们在应力分布上的影响, 对时间点进行加权. 去看看空间强度测度的重分形特性的改变是否是一个先导, 或是否与所观测的地震检测方式的改变相吻合, 是非常有趣的.

在每一种所分析的情形中, Rényi 维数 D_q 的估计随 q 的增长而递减. 如果接受地震的空间强度测度支撑在一个自相似集上, 则 Rényi 维数的这种递减暗示了在这个地区的某个地方, 地震会有更大的可能出现. 如果空间强度测度是一致分布的, 则希望所有的 Rényi 维数相同.

11.5.2 相关研究的汇总

有很多文章给出了地震位置的维数估计. Hirata 和 Imoto(1991) 估计了关东地区震源的 Rényi 维数的一个值, 给出关联维数大约为 2.2. Geilikman 等 (1990) 也用了一个重分形的方法研究了帕米尔、高加索以及加利福尼亚地区地震震源的空间分布, 他们用计盒的方法估计了 Rényi 维数, 得到了上述几个地区的盒维数 $\widetilde{D}_0$ 大约为 2.0.

Eneva(1996) 对于采矿引起的地震活动的空间分布估计了不同的 Rényi 维数, 她的方法是去计算中心位于数据集中, 以某个值为半径的球中地震的数目. 她推断对于有限的数据集, 明显的重分形行为是不可能的.

Kagan(1991) 用了一个全球范围的地震分类去估计震中和震源的分形维数. 他发现, 当时间增长时, 浅地震的震源的维数估计值渐渐到达 2.1~2.2, 而中等深度地震 (71~280 公里) 的维数估计下降到 1.8~1.9, 深地震则下降到 1.5~1.6. 地震越深, 维数越小, 与这一章的结果正好相反.

Kagan(1991) 所考虑的点间距比我们考虑的大得多. 他的方法是去考察一个全球范围的震中的地图, 将会看到浅地震都集中在地震带中, 这个地震带连续地、光

滑地沿着环太平洋延伸, 而深地震都集中在几个俯冲区域中. 这暗示了深地震的维数估计比浅地震的维数估计值更小, 至少对于距离大于 50 公里的情形是这样的. 我们分析的是更小的局部区域, 两个地区都在俯冲区域中. 一个集合的维数是当距离趋于零时, 定义为某个极限. 然而, 当考虑的集合是自相似的, 则在距离的某个范围内就会有某种幂率变化, 幂率指数与维数相同. 为了用更大的点间距, 必须做两个假设, 第一是边界调整的合法性, 第二是大距离的幂率指数与小距离的幂率指数相同. 已经分析了小的点间距, 结果不得不与震源的位置误差相抗衡. 我们分析中所使用的尺度大约等价于地区地震模型中所使用的尺度.

震源的位置误差可以用来解释在深地震的情形, 我们的维数估计值比 Kagan (1991) 的要大, 虽然震中的维数估计不是这样的. 而在浅地震的情形, 我们却得到比 Kagan(1991) 的更小的值, 这与位置误差的影响也不一致. 这种差别可能是由于我们采用相对短的时间周期所引起. Kagan(1991) 指出, 当分类中时间周期增长时, 维数会渐渐递增到 D_2. 然而, 在关东地区和惠灵顿地区, 采用的是同样的时间长度. Ogata 和 Abe(1991) 提出, 地震活动表现出的是长期变化, 因此, 相对于较少的地震活动, 更多的地震活动所在的时间周期中会出现不同的空间模型.

第四部分

附　　录

附录 A　集合的性质和维数

在本附录中, 给出了盒维数、Hausdorff 维数和 Packing 维数的定义, 以及它们之间的各种关系. 大部分的结果的表述来自 Falconer (1990) 这一优秀著作. 在此仅给出一些主要结果的汇总, 更多的细节可以在 Falconer (1990) 中找到. 更多的数学上的细节也可以参考 Falconer (1985). Tricot (1995) 一书也讨论了各种维数概念, 但此书更多的关注曲线.

A.1　自相似集

A.1.1　定义 (Falconer, 1990, 第 9 章)

1. 令 E 为 $\mathbb{R}^d$ 的一个闭子集. 称映射 $S\colon E \to E$ 为 E 上的**压缩映射**, 若 $\exists t\colon 0 < t < 1\colon |S(x) - S(y)| \leqslant t|x - y|, \forall x, y \in E$.

2. 令 E 为 $\mathbb{R}^d$ 的一个闭子集. 称映射 $S\colon E \to E$ 为 E 上的**相似映射**, 若 $\exists t\colon 0 < t < 1\colon |S(x) - S(y)| = t|x - y|, \forall x, y \in E$. 即在几何上, S 将一个集合变换成一个与之相似的集合.

3. 令 $S_1, \cdots, S_m$ 为压缩映射. 称 E 的子集 F 在变换 S_i 下**不变**, 若

$$F = \bigcup_{i=1}^{m} S_i(F). \tag{A.1}$$

4. 称一个在一族相似映射下不变的集合为**自相似集**, 即它能表示为一些自身的小拷贝的并.

5. 称压缩映射 $\{S_i\}$ 满足**开集条件**, 若存在非空有界开集 V, 使得

$$V \supseteq \bigcup_{i=1}^{m} S_i(V), \tag{A.2}$$

并且这个并是不交并.

请注意开集条件和集合互不相交条件的区别. 在一些文献中, 开集条件可能被描述为**不重叠**; 即, 它们只相交于边界上的点.

A.1.2　定理 (Falconer, 1990, 114 页)

令 $S_1, \cdots, S_m$ 为 $\mathbb{R}^d$ 子集 E 上的压缩映射, 满足

$$|S_i(x) - S_i(y)| \leqslant t_i|x - y|, \quad \forall x, y \in E,$$

其中, 对于每个 i, 都有 $t_i < 1$. 则存在唯一的非空紧集 F, 它在 S_i 下不变, 即满足式 A.1. 注意到 $F \subseteq \mathbb{R}^d$ 为紧集, 所以它一定是闭的.

A.1.3 例

设 F 为 $[0,1]$ 区间上的 Cantor 三分集. 令 $S_1(x) = x/3$, $S_2(x) = x/3 + 2/3$. 则 $F = S_1(F) \cup S_2(F)$, F 是自相似的. 进一步, 令 $V = (0,1)$, 则 $S_1(V) = (0, \frac{1}{3})$, $S_2(V) = (\frac{2}{3}, 1)$, 所以 $\{S_1, S_2\}$ 满足开集条件.

A.1.4 定义 (Falconer, 1990, 第 9 章)

1. **仿射映射** $A: \mathbb{R}^d \to \mathbb{R}^d$ 指具有 $A(x) = T(x) + b$ 形式的映射, 其中 $x \in \mathbb{R}^d$, T 为一个非奇异的线性变换.

2. 令 $A_1, \cdots, A_m$ 为 $\mathbb{R}^d$ 上的仿射压缩映射. 若集合 F 满足

$$F = \bigcup_{i=1}^{m} A_i F,$$

则称 F 是**自仿**的.

A.1.5 定理 (Falconer, 1990, 118 页)

设 S_i 为 $\mathbb{R}^d$ 上压缩率为 t_i $(i = 1, \cdots, m)$ 的相似映射, 并且满足开集条件 (式 (A.2)). 若 F 为满足式 (A.1) 的不变集, 则 $\dim_H F = \dim_B F = s$, 其中 s 满足

$$\sum_{i=1}^{m} t_i^s = 1.$$

并且, 对于这样的 s, $0 < \mathcal{H}^s(F) < \infty$.

对于 S_i 为压缩映射, 但不是相似映射的情形, Falconer (1990, §9.3) 给出了进一步的结果.

A.2 Hausdorff 维数

更详细的数学上的处理见 Falconer (1985, §1.2), Falconer (1990, §2.1) 和 Rogers (1970).

令 U 为 $\mathbb{R}^d$ 的非空子集. U 的直径定义为

$$|U| = \sup\{|x - y| : x, y \in U\}.$$

若 $F \subset \bigcup_{i=1}^{\infty} U_i$, 并且对于每个 i, 有 $0 < |U_i| \leqslant \delta$, 则称 $\{U_i\}$ 为 F 的 **δ-覆盖**.

A.2.1 Hausdorff 测度的定义 (Falconer, 1990, §2.1)

设 F 为 $\mathbb{R}^d$ 的子集, s 为一个非负数. 对于任意 $\delta > 0$, 定义 $\mathcal{H}^s_\delta(F)$ 为

$$\mathcal{H}^s_\delta(F) = \inf\left\{\sum_{i=1}^{\infty} |U_i|^s : \{U_i\} \text{ 为 } F \text{ 的 } \delta\text{-覆盖}\right\},$$

其中下确界取遍所有的 F 的可数 δ-覆盖. F 的 s-维 Hausdorff 测度定义为

$$\mathcal{H}^s(F) = \lim_{\delta \to 0} \mathcal{H}^s_\delta(F).$$

极限 $\mathcal{H}^s(F)$ 对于所有的 $F \subset \mathbb{R}^d$ 都存在, 尽管这个极限值常常是 0 或 ∞.

A.2.2 Hausdorff 测度的性质 (Falconer, 1990, §2.1)

下面列出 Hausdorff 测度的各种性质.

1. (a) $\mathcal{H}^s(\varnothing) = 0$.

(b) 若 $E \subseteq F$, 则 $\mathcal{H}^s(E) \leqslant \mathcal{H}^s(F)$.

(c) 若 $\{F_i\}$ 为一族可数的互不相交的 Borel 集, 则

$$\mathcal{H}^s\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} \mathcal{H}^s(F_i).$$

2. 若 F 为 $\mathbb{R}^d$ 的 Borel 子集, 则

$$\mathcal{H}^d(F) = c_d \operatorname{Vol}^d(F),$$

其中 $c_d = \pi^{d/2} 2^d (d/2)!$. 例如, 若 F 为光滑曲面, 则

$$\mathcal{H}^2(F) = \frac{\pi}{4} \operatorname{Area}(F).$$

3. 比例性质. 若 $F \subset \mathbb{R}^d$, $\lambda > 0$, 则 $\mathcal{H}^s(\lambda F) = \lambda^s \mathcal{H}^s(F)$, 其中 $\lambda F = \{\lambda x : x \in F\}$.

4. 令 $F \subset \mathbb{R}^d$, $f : F \to \mathbb{R}^d$ 为满足下面条件的映射,

$$|f(x) - f(y)| \leqslant c|x - y|^\alpha,$$

其中 $x, y \in F$, 常数 $c > 0$, $\alpha > 0$. 则对于每个 s,

$$\mathcal{H}^{s/\alpha}(f(F)) \leqslant c^{s/\alpha} \mathcal{H}^s(F).$$

若 $t > s$, $\{U_i\}$ 为 F 的 δ-覆盖, 则

$$\sum_{i=1}^{\infty} |U_i|^t \leqslant \delta^{t-s} \sum_{i=1}^{\infty} |U_i|^s,$$

再对所有可能的覆盖取下确界,

$$\mathcal{H}_\delta^t(F) \leqslant \delta^{t-s}\mathcal{H}_\delta^s(F).$$

令 $\delta \to 0$, 可以看到, 若 $\mathcal{H}^s(F) < \infty$, 则 $\mathcal{H}^t(F) = 0$ 对所有的 $t > s$. 所以, 在以 s 为自变量的 $\mathcal{H}^s(F)$ 的图像上存在一个临界值, 在这个临界值附近, $\mathcal{H}^s(F)$ 的值从无穷跳跃到零.

A.2.3　Hausdorff 维数 (Falconer, 1990, §2.2)

集合 F 的 Hausdorff 维数, 记为 $\dim_H(F)$, 定义为

$$\dim_H(F) = \inf\{s\colon \mathcal{H}^s(F) = 0\} = \sup\{s\colon \mathcal{H}^s(F) = 0\},$$

或者等价地, 满足下面的条件

$$\mathcal{H}^s(F) = \begin{cases} \infty, & 若\ s < \dim_H(F), \\ 0, & 若\ s > \dim_H(F). \end{cases}$$

若 $s = \dim_H(F)$, 则 $\mathcal{H}^s(F)$ 可能为零、无穷或 $0 < \mathcal{H}^s(F) < \infty$. 满足最后这个条件的 Borel 集被称为 s-集.

A.2.4　Hausdorff 维数的性质 (Falconer, 1990, §2.2)

Hausdorff 维数满足下面的性质.

1. 若 $F \subset \mathbb{R}^d$ 为开集, 则 $\dim_H(F) = d$.
2. 若 F 为 $\mathbb{R}^d$ 中的 m-维连续光滑子流形, 则 $\dim_H(F) = m$.
3. 若 $E \subset F$, 则 $\dim_H(E) \leqslant \dim_H(F)$.
4. 若 $F_1, F_2, \cdots$ 为可数多个集合的序列, 则

$$\dim_H\left(\bigcup_{i=1}^{\infty} F_i\right) = \sup_{1\leqslant i<\infty} \dim_H(F_i).$$

5. 若 F 可数, 则 $\dim_H(F) = 0$.
6. 若 $F \subset \mathbb{R}^d$ 与 $E \subset \mathbb{R}^m$ $(m \leqslant d)$ 是等距同构的, 则 $\dim_H(F) = \dim_H(E)$. 若 E 有正的 m-维 Lebesgue 测度, 则

$$\dim_H(F) = \dim_H(E) = m.$$

A.2.5　更多的定义 (Falconer, 1990, §2.5)

设 $h\colon \mathbb{R}^+ \to \mathbb{R}^+$ 为一个连续的单调上升的函数. 对于 $\mathbb{R}^d$ 的子集 F, 令

$$\mathcal{H}_\delta^h(F) = \inf\left\{\sum_{i=1}^{\infty} h(|U_i|)\colon \{U_i\}\ 为\ F\ 的一个\ \delta\text{-覆盖}\right\}.$$

由此可诱导一个测度, 令

$$\mathcal{H}^h(F) = \lim_{\delta\to 0} \mathcal{H}^h_\delta(F).$$

函数 h 被称为维数函数.

s-维 Hausdorff 测度的定义中所用到的维数函数 $h(t) = t^s$. 若 h 和 g 为维数函数, 满足

$$\lim_{t\to 0} \frac{h(t)}{g(t)} = 0,$$

则可以证明, 只要 $\mathcal{H}^g(F) < \infty$, 就有 $\mathcal{H}^h(F) = 0$. 例如, 一个具有维数函数 $g(t) = t^s|\log t|$ 的集合可能 Hausdorff 维数为 s, 尽管这个集合可能比一个具有维数函数 $h(t) = t^s$ 的集合稍小.

S.J. Taylor 参与合作了多篇关于各种随机过程的维数函数的文章. Taylor (1986) 给出了一个近期的评论.

A.3 盒 维 数

Falconer (1990, §3.1) 包含了本节的内容, 并提供了更多的细节.

A.3.1 定义 (Falconer, 1990, 38 页)

记 $N_\delta(F)$ 为覆盖 $\mathbb{R}^d$ 的非空有界子集 F 所需的直径至多为 δ 的集合的最少个数. F 的上盒维数和下盒维数的定义分别为

$$\underline{\dim}_B(F) = \liminf_{\delta\to 0} \frac{\log N_\delta(F)}{-\log\delta}$$

和

$$\overline{\dim}_B(F) = \limsup_{\delta\to 0} \frac{\log N_\delta(F)}{-\log\delta}.$$

若 $\underline{\dim}_B(F) = \overline{\dim}_B(F)$, 则 F 的盒维数定义为

$$\dim_B(F) = \lim_{\delta\to 0} \frac{\log N_\delta(F)}{-\log\delta}.$$

A.3.2 定理 (Falconer, 1990, 40 页)

$N_\delta(F)$ 可以定义为下列任何一个数, 对于盒维数的定义来说, 它们都是等价的.

1. 覆盖 F 所需的半径为 δ 的闭球的最少个数.
2. 覆盖 F 所需的边长为 δ 的立方体的最少个数.
3. 覆盖 F 所需的边长为 δ 的立方体网格的最少个数.
4. 覆盖 F 所需的直径至多为 δ 的集合的最少个数.
5. 球心在 F 上的互不相交的球的最大个数.

A.3.3　注

$\underline{\dim}_B(F)$ 可以用相等半径的小球所形成的最经济的覆盖来定义 (即上面的 (1)). $\overline{\dim}_B(F)$ 可以认为是这样的一个维数, 它依赖于尽可能密集的相同半径的互不相交的小球所形成的覆盖 (即上面的 §A.3.2 中的 5). 这些概念是后面定义的 Packing 测度的基础.

A.3.4　命题 (Falconer, 1990, 44 页)

用 $\overline{F}$ 表示 F 的闭包 (即包含 F 的 $\mathbb{R}^d$ 的最小闭子集). 则

$$\underline{\dim}_B(\overline{F}) = \underline{\dim}_B(F)$$

以及

$$\overline{\dim}_B(\overline{F}) = \overline{\dim}_B(F).$$

A.3.5　例

下面的例子是一些与盒维数相关的问题.

1. 令 F 为 $(0,1)$ 上有理数组成的集合 (可数集). 则 $\overline{F} = [0,1]$, 所以 $\underline{\dim}_B(F) = \overline{\dim}_B(F) = 1$, 并且 $\dim_B(F) = \dim_B(\overline{F}) = 1$. 所以可数集也可能有非零的盒维数.

2. 令 $F = \{0, 1, \frac{1}{2}, \frac{1}{3}, \cdots\}$, 则 F 为紧集, 并且 $\dim_B(F) = \frac{1}{2}$ (Falconer, 1990, 45 页).

A.3.6　与 Hausdorff 维数的比较

注意到

$$N_\delta(F)\delta^s \longrightarrow \begin{cases} \infty, & \text{若 } s < \dim_B(F), \\ 0, & \text{若 } s > \dim_B(F). \end{cases}$$

又

$$N_\delta(F)\delta^s = \inf\left\{\sum_i \delta^s \colon \{U_i\} \text{ 为 } F \text{ 的一个有限的 } \delta\text{-覆盖}\right\},$$

以及

$$\mathcal{H}^s_\delta(F) = \inf\left\{\sum_{i=1}^\infty |U_i|^s \colon \{U_i\} \text{ 为 } F \text{ 一个 } \delta\text{-覆盖}\right\}.$$

与 Hausdorff 维数的定义相比较, 自然会考虑将

$$\inf\left\{s\colon \liminf_{\delta\to 0} N_\delta(F)\delta^s = 0\right\}$$

作为盒维数的定义. 然而, $\liminf_{\delta\to 0} N_\delta(F)\delta^s$ 并不是一个定义在 $\mathbb{R}^d$ 的子集上的测度. 所以, 与 Hausdorff 维数不同, 盒维数并不是通过测度来定义的. 但是, Pesin (1993, §2.2) 定义了上、下 s-盒测度.

盒维数可以认为是这样的一个指标, 它表明了用相同大小的小集合去覆盖一个集合的效率, 而 Hausdorff 维数定义中所涉及的覆盖, 所用到的小集合其大小可能是相差很大的.

A.4 Packing 维 数

Packing 维数是由 Tricot (1982) 引入的. 虽然用到了 $\overline{\dim}_B$ 的一些概念 (见注 A.3.3), 但它是在一个测度的基础上, 用与 Hausdorff 维数类似的方式定义的. 令 $F \subset \mathbb{R}^d$, 定义

$$\mathcal{P}_\delta^s(F) = \sup\left\{\sum_{i=1}^{\infty}|B_i|^s \colon \{B_i\}\ \text{为一族球心在}\ F\ \text{且半径至多为}\ \delta\ \text{的互不相交的球}\right\}.$$

因为 $\mathcal{P}_\delta^s(F)$ 随 δ 单调下降, 极限

$$\mathcal{P}_0^s(F) = \lim_{\delta\to 0}\mathcal{P}_\delta^s(F)$$

存在. 然而, $\mathcal{P}_0^s(F)$ 并不是一个测度, 比如当 F 为一个可数稠密集时. $\mathcal{P}_0^s$ 没有可加性, 所以, 需要作进一步的修正.

A.4.1 Packing 测度的定义 (Falconer, 1990, 47 页)

$\mathbb{R}^d$ 上子集 F 的 s-维 Packing 测度定义为

$$\mathcal{P}^s(F) = \inf\left\{\sum_{i=1}^{\infty}\mathcal{P}_0^s(F_i) \colon F \subset \bigcup_{i=1}^{\infty} F_i\right\}.$$

A.4.2 Packing 维数的定义 (Falconer, 1990, 47 页)

Packing 维数 $\dim_P(F)$ 定义为

$$\dim_P(F) = \sup\{s \colon \mathcal{P}^s(F) = \infty\} = \inf\{s \colon \mathcal{P}^s(F) = 0\}.$$

A.4.3 注 (Falconer, 1990, 47 页)

若对所有的 i, $s > \dim_P(F_i)$, 则

$$\mathcal{P}^s\left(\bigcup_{i=1}^{\infty} F_i\right) \leqslant \sum_{i=1}^{\infty}\mathcal{P}^s(F_i) = 0.$$

所以, $\dim_P\left(\bigcup_{i=1}^{\infty} F_i\right) \leqslant s$. 同样注意到

$$\dim_P\left(\bigcup_{i=1}^{\infty} F_i\right) = \sup_i \dim_P(F_i).$$

A.4.4　引理 (Falconer, 1990, 43, 48 页)

对于 $F \subset \mathbb{R}^d$, 下面的关系成立:

$$\dim_H(F) \leqslant \dim_P(F) \leqslant \overline{\dim}_B(F),$$

以及

$$\dim_H(F) \leqslant \underline{\dim}_B(F) \leqslant \overline{\dim}_B(F).$$

附录B 大 偏 差

B.1 导 论

在这一章, 给出与大偏差相关的一些结果的概述, 第 4 ~ 6 章的内容涉及到这些结果; 尤其是 Gärtner-Ellis 定理. 大部分的定义和结果来自于 Ellis (1995) 一书, 在这本书中还能找到相当高深的内容. Bucklew (1990) 一书提供了这门学科容易理解的入门介绍.

考虑一列随机变量 $Y_1, Y_2, \cdots$, 它依概率收敛于一个实常数 y_0. 即

$$\Pr\{|Y_n - y_0| > \varepsilon\} \to 0, \quad n \to \infty.$$

常常有这样的情况, $\Pr\{|Y_n - y_0| > \varepsilon\}$ 不仅趋向于 0, 并且是以指数速度收敛. 我们将更清楚的描述这样的收敛, 为此考虑与 y 有关的 ε 区间, $(y-\varepsilon, y+\varepsilon)$, 它不一定包含极限点 y_0; 然后考察当 $n \to \infty$ 时的概率 $\Pr\{Y_n \in (y-\varepsilon, y+\varepsilon)\}$. 常常会有

$$\Pr\{Y_n \in (y-\varepsilon, y+\varepsilon)\} \approx L(\varepsilon, y, n)\exp[-nI_0(y, \varepsilon)],$$

其中 $L(\varepsilon, y, n)$ 非负, $\log L(\varepsilon, y, n) = o(n)$ 当 $n \to \infty$, 并且 $I_0(y, \varepsilon)$ 为一个非负函数, 当 $y = y_0$ 时, 取值为零. 对于固定的 $\varepsilon > 0$ 以及很大的 n, 若 $|Y_n - y_0| > \varepsilon$, 则 Y_n 可以认为是偏离正常值 y_0 的一个*大偏差*, 这是该术语的由来. 现在将上式重新表示为

$$\frac{\log\Pr\{Y_n \in (y-\varepsilon, y+\varepsilon)\}}{n} \approx \frac{\log L(\varepsilon, y, n)}{n} - I_0(y, \varepsilon).$$

因为 $\log L(\varepsilon, y, n) = o(n)$, 当 $n \to \infty$, 则对于固定的 ε 和充分大的 n,

$$\frac{\log\Pr\{Y_n \in (y-\varepsilon, y+\varepsilon)\}}{n} \approx -I_0(y, \varepsilon).$$

在相当广泛的条件下, 这可表示为

$$\lim_{n\to\infty}\frac{1}{a_n}\log\Pr\{Y_n \in A\} = -\inf_{y\in A} I(y), \tag{B.1}$$

其中, $\{a_n; n = 1, 2, \cdots\}$ 为一列趋于无穷的正数, $A \in \mathcal{B}(\mathbb{R})$, 并且 $I(y)$ 为一个非负凸函数, 满足 $I(y_0) = 0$. 大偏差理论的主要内容是: 在什么条件下, 上述或类似的结论成立, 以及确定函数 $I(y)$ 的表达式.

设 Y_n 为一列独立同分布随机变量的经验平均, 在这种情况下, Cramér 定理证明函数 $I(y)$ 与累积生成函数相关. 下一节将介绍 Cramér 定理.

Eills (1984, 1985) 和 Gärtner (1977) 考虑了更一般的情况, 其中 Y_n 在 $\mathbb{R}^k$ 中取值, 不一定为经验平均, 并且可以定义在一列不同的概率空间 $(\Omega_n, \mathcal{F}_n, \rho_n)$ 上. 如果 Y_n 定义在同一概率空间 $(\Omega, \mathcal{F}, \rho)$ 上, 并且式 (B.1) 中的序列 a_n 增长的足够快, 则 $Y_n \to y_0$, ρ-a.s. §B.3 将介绍这些结果.

B.2　Cramér　定　理

最早的结果之一由 Harald Cramér 给出. 他的结果要求累积生成函数是陡峭的, 下面马上给出定义.

B.2.1　定义 (Ellis, 1984)

记函数 f 的定义域为 $\mathcal{D}(f) = \{x\colon f(x) < \infty\}$. 设函数 $f\colon \mathbb{R}^k \to \mathbb{R}$ 在定义域 $\mathcal{D}(f)$ 的内部可微, 称 f 是陡峭的, 若对于任意趋于 $\mathcal{D}(f)$ 的某个边界点的点列 $\{x_n\} \subseteq \mathcal{D}(f)$, 都有

$$\|\nabla f(x_n)\| = \left\|\left(\frac{\partial f(x_n)}{\partial x_{n1}}, \cdots, \frac{\partial f(x_n)}{\partial x_{nd}}\right)\right\| \to \infty.$$

B.2.2　Cramér 定理 (Bucklew, 1990, 7 页)

令 $\{X_i\}, i = 1, 2, \cdots$ 为一列独立同分布随机变量, 并且

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{B.2}$$

设 $E[X_1]$ 存在且有限. 并且, 定义函数 $I(y)$ 为

$$I(y) = \sup_q \{qy - C(q)\}, \quad y \in \mathbb{R},$$

其中 $C(q)$ 为 X_1 的累积生成函数, 为

$$C(q) = \log E\left[e^{qX_1}\right].$$

设 $C(q)$ 是陡峭的. 则对于每个闭集 $K \subseteq \mathbb{R}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Y_n \in K\} \leqslant -\inf_{y\in K} I(y);$$

并且对于每个开集 $G \subseteq \mathbb{R}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Y_n \in G\} \geqslant -\inf_{y\in G} I(y).$$

B.2.3 推论

若对 $a, b \in \mathbb{R}$ 且 $a < b$,

$$\inf_{y\in[a,b]} I(y) = \inf_{y\in(a,b)} I(y),$$

则

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\{Y_n \in (a,b)\} = -\inf_{y\in(a,b)} I(y).$$

B.2.4 例

令 $\{X_i\}$, $i = 1, 2, \cdots$ 为一列独立同分布 Gauss 随机变量, 具有平均值 μ 和单位方差. 则累积生成函数为

$$C(q) = \log E\left[e^{qX_1}\right] = \frac{q^2}{2} + \mu q, \quad q \in \mathbb{R},$$

并且

$$I(y) = \frac{1}{2}(y-\mu)^2.$$

因 $q \in \mathbb{R}$, 所以对于 $C(q)$ 的定义域, 不存在边界问题. 记 Y_n 为式 (B.2) 中的经验平均, 则由 Cramér 定理可得

$$\lim_{n\to\infty} \frac{\log \Pr\{Y_n \in A\}}{n} = -\inf_{z\in A} I(z), \quad A \in \mathcal{B}(\mathbb{R}).$$

考虑区间 A 的各种可能性.

1. 令 $A = (\mu-\varepsilon, \mu+\varepsilon)$. 则 $\inf_{z\in A} I(z) = 0$, 因此

$$\lim_{n\to\infty} \frac{\log \Pr\{|Y_n - \mu| < \varepsilon\}}{n} = 0.$$

2. 令 $A = (-\infty, \mu-\varepsilon) \cup (\mu+\varepsilon, \infty)$. 则 $\inf_{z\in A} I(z) = \varepsilon^2/2$, 因此

$$\lim_{n\to\infty} \frac{\log \Pr\{|Y_n - \mu| > \varepsilon\}}{n} = -\frac{\varepsilon^2}{2}.$$

另一种方法, $Y_n \sim \mathcal{N}(\mu, 1/n)$, 所以 $\Pr\{|Y_n - \mu| > \varepsilon\} = 2\Pr\{Z > \sqrt{n}\varepsilon\}$, 其中 $Z \sim \mathcal{N}(0,1)$. 由 Abramowitz 和 Stegun (1964, 式 (26.2.14)) 可得, 对于 $\varepsilon > 0$

$$\Pr\{|Y_n - \mu| > \varepsilon\} = \left\{ \frac{1}{\sqrt{n}\varepsilon+} \cdot \frac{1}{\sqrt{n}\varepsilon+} \cdot \frac{2}{\sqrt{n}\varepsilon+} \cdot \frac{3}{\sqrt{n}\varepsilon+} \cdot \frac{4}{\sqrt{n}\varepsilon+} \cdots \right\} \frac{2}{\sqrt{2\pi}} \exp\left(\frac{-n\varepsilon^2}{2}\right).$$

大括号中的项为连分数. 在两边取对数, 除以 n, 再对 n 取极限, 就得到同样的结果.

3. 现在考虑一个关于点 y 的 ε 区间, 这里 $y \neq \mu$, 即 $A = (y-\varepsilon, y+\varepsilon)$, 并且 ε 充分小, 使得 $\mu \notin A$. 则

$$\inf_{z \in A} I(z) = \frac{1}{2}(|y-\mu| - \varepsilon)^2,$$

从而

$$\lim_{n\to\infty} \frac{\log \Pr\{|Y_n - y| < \varepsilon\}}{n} = -\frac{1}{2}(|y-\mu| - \varepsilon)^2.$$

在两边对 ε 取极限得到

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} \frac{\log \Pr\{|Y_n - y| < \varepsilon\}}{n} = -I(y).$$

这个式子刻画了概率函数的局部收敛速率.

B.2.5 例

令 $\{X_i\}, i = 1, 2, \cdots$ 为一列独立同分布指数随机变量, 其累积生成函数为

$$C(q) = \log E\left[e^{qX_1}\right] = -\log\left(1 - \frac{q}{\lambda}\right), \quad \lambda > 0 \text{ 且 } -\infty < q < \lambda,$$

因此, $I(y) = \lambda y - 1 - \log(\lambda y)$. 注意到 $C(q)$ 的定义域边界为 $q = \lambda$, 然而可以看到, 当 $q \nearrow \lambda$, $C(q)$ 是陡峭的. 记 Y_n 为式 B.2 中的经验平均, 则由 Cramér 定理得到, 对于 $y > 1/\lambda$,

$$\lim_{n\to\infty} \frac{\log \Pr\{Y_n > y\}}{n} = -\inf_{z\in(y,\infty)} I(z) = 1 + \log(\lambda y) - \lambda y.$$

B.2.6 例

令 $\{X_i\}, i = 1, 2, \cdots$ 为一列独立同分布标准 Gauss 随机变量, 考虑部分和 Y_n, 其中

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

X_1^2 具有 χ^2 平方分布, 其累积生成函数为

$$C(q) = \log E\left[e^{qX_1^2}\right] = \frac{-\log[1-2q]}{2}, \quad -\infty < q < \frac{1}{2},$$

因此 $I(y) = \frac{1}{2}(y - \log y - 1)$. 可以看到 $C(q)$ 在边界 $q = \frac{1}{2}$ 处是陡峭的, 从而由 Cramér 定理可得, 对于 $y > 1$,

$$\lim_{n\to\infty} \frac{\log \Pr\{Y_n > y\}}{n} = -\inf_{z\in(y,\infty)} I(z) = \frac{-y + \log y + 1}{2}.$$

B.3 Gärtner-Ellis 定理

这一节将结果推广到多重维数、非独立同分布、更一般的比例常数, 以及不一定与部分和有关的情形. 这些结果来自 Ellis(1984, 1985) 与 Gärtner (1977) 中的结果类似.

B.3.1 定义 (Ellis, 1985, 第 6 章)

1. 称集合 $A \subseteq \mathbb{R}^k$ 为凸的, 若对所有 $y_1, y_2 \in A$ 及所有 $0 < \lambda < 1$, 都有 $\lambda y_1 + (1 - \lambda y_2) \in A$.

2. 令 f 为定义在 $A \subseteq \mathbb{R}^k$ 上取广义实值的函数. 称 *f* *在* *A* *上凸*, 若

(a) A 是凸的,

(b) 至少存在一个 $y \in A$, 使得 $f(y)$ 是有限的,

(c) f 不能取 $-\infty$,

(d) 对于所有 $y_1, y_2 \in A$ 以及所有 $0 < \lambda < 1$, 都有 $f(\lambda y_1 + (1-\lambda)y_2) \leqslant \lambda f(y_1) + (1-\lambda) f(y_2)$.

3. 令 f 为 $\mathbb{R}^k$ 上的凸函数. y_2 为 $\mathbb{R}^k$ 中的一点, 称向量 $z \in \mathbb{R}^k$ 为 *f* *在* y_2 *的次梯度* (见 §B.3.2 中的解释), 若

$$f(y_1) \geqslant f(y_2) + \langle z, y_1 - y_2 \rangle, \quad \text{对于所有 } y_1 \in \mathbb{R}^k.$$

4. *f* *在* *y* *的次微分*定义为集合

$$\partial f(y) = \left\{ z \in \mathbb{R}^k \colon z \text{ 为 } f \text{ 在 } y \text{ 的次梯度} \right\}.$$

B.3.2 注

在 f 的不可导点处, 次梯度的概念推广了导数的概念. 由上可知, 若 f 凸, 则

$$\lambda(f(y_1) - f(y_2)) \geqslant f(y_2 + \lambda(y_1 - y_2)) - f(y_2).$$

若 f 在 y_2 可微, 则当 $\lambda \to 0^+$ 时,

$$\lambda(f(y_1) - f(y_2)) \geqslant \lambda \langle \nabla f(y_2), y_1 - y_2 \rangle + o(\lambda \| y_1 - y_2 \|).$$

重组上式, 并且令 $\lambda \to 0$, 得到

$$f(y_1) \geqslant f(y_2) + \langle \nabla f(y_2), y_1 - y_2 \rangle.$$

B.3.3　预备知识

令 $\{U_n; n = 1, 2, \cdots\}$ 为定义在概率空间 $\{(\Omega_n, \mathcal{F}_n, \rho_n); n = 1, 2, \cdots\}$ 上的一列随机向量, 即

$$U_n \colon (\Omega_n, \mathcal{F}_n) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k)),$$

其中 $\mathcal{B}(\mathbb{R}^k)$ 为 $\mathbb{R}^k$ 中的 Borel 集族. 同样考虑重比例化后的随机序列 $\{Y_n; n = 1, 2, \cdots\}$, 这里

$$Y_n = \frac{U_n}{a_n},$$

并且 $\{a_n; n = 1, 2, \cdots\}$ 为一列趋于 ∞ 的正数序列. 令 $\{Q_n; n = 1, 2, \cdots\}$ 为 $\mathcal{B}(\mathbb{R}^k)$ 上的一列概率测度, 满足

$$Q_n(B) = \rho_n\{\omega \in \Omega_n \colon Y_n(\omega) \in B\},$$

其中 $B \in \mathcal{B}(\mathbb{R}^k)$.

B.3.4　定义 (Ellis, 1985, 35 页)

令 $\{Q_n; n = 1, 2, \cdots\}$ 为 $\mathcal{B}(\mathbb{R}^k)$ 上的一列概率测度, 称 $\{Q_n\}$ 具有*大偏差性质*, 若存在趋于 ∞ 的正数序列 $\{a_n; n = 1, 2, \cdots\}$, 及将 $\mathbb{R}^k$ 映到 $[0, \infty]$ 的函数 $I(y)$, 满足下面的假设.

1. $I(y)$ 在 $\mathbb{R}^k$ 上是闭的 (下半连续, 即 $y_n \to y \Rightarrow \liminf_{n\to\infty} I(y_n) \geqslant I(y)$).
2. $I(y)$ 具有紧的水平集, 即对于每个 $b \in \mathbb{R}$, $\{y \in \mathbb{R}^k \colon I(y) \leqslant b\}$ 是紧的.
3. 对于每个 $\mathbb{R}^k$ 中的闭集 K,

$$\limsup_{n\to\infty} \frac{1}{a_n} \log Q_n\{K\} \leqslant -\inf_{y\in K} I(y). \tag{B.3}$$

4. 对于每个 $\mathbb{R}^k$ 中的开集 G,

$$\liminf_{n\to\infty} \frac{1}{a_n} \log Q_n\{G\} \geqslant -\inf_{y\in G} I(y). \tag{B.4}$$

称 $I(y)$ 为 $\{Q_n\}$ 的*熵函数*.

B.3.5　定理 (Ellis, 1985, 36 页)

若对于一个固定的序列 $\{a_n; n = 1, 2, \cdots\}$, $\{Q_n\}$ 对于熵函数 I 和 J 具有大偏差性质, 则对于所有的 $y \in \mathbb{R}^k$, $I(y) = J(y)$ (即唯一性).

B.3.6 定理 (Ellis, 1985, 37 页)

令 $\{Q_n\}$ 为 $Y_n = U_n/a_n$ 在 $\mathbb{R}^k$ 上的分布. 设 Q_n 对于常数 $\{a_n; n = 1, 2, \cdots\}$ 及熵函数 $I(y)$ 具有大偏差性质. 称 Borel 集 A 为 I-连续集, 若

$$\inf_{y \in \mathrm{cl}\, A} I(y) = \inf_{y \in \mathrm{int}\, A} I(y).$$

若 A 为 I-连续集, 则

$$\lim_{n \to \infty} \frac{1}{a_n} \log Q_n\{A\} = -\inf_{y \in A} I(y).$$

B.3.7 重比例化累积生成函数

定义函数 $C_n(q)$ 为

$$\begin{aligned} C_n(q) &= \frac{1}{a_n} \log E_n[\exp\langle q, U_n \rangle] \\ &= \frac{1}{a_n} \log \int_{\Omega_n} \exp\langle q, U_n(\omega) \rangle \rho_n(d\omega), \end{aligned} \tag{B.5}$$

其中 $\{a_n; n = 1, 2, \cdots\}$ 为一列趋于 ∞ 的正实数, 且 $q \in \mathbb{R}^k$. 若 $\lim_{n\to\infty} C_n(q)$ 存在, 则称 $C(q) = \lim_{n\to\infty} C_n(q)$ 为 $\{U_n\}$ 的**重比例化累积生成函数**. 值得注意的是 $C(q)$ 也常常被称为**自由熵函数**.

B.3.8 假设 (Ellis, 1984, 假设 II.1)

设下面的假设成立.

1. 对于所有的 $q \in \mathbb{R}^k$, $C(q) = \lim_{n\to\infty} C_n(q)$ 存在, 这里允许 $+\infty$ 同时作为极限值和序列 $C_n(q)$ 中的元 (即定义 $C(q) = \infty$, 若对于所有 $n > n_0$, $C_n(q) = \infty$, n_0 依赖于 q).
2. $\mathcal{D}(C)$ 具有包含点 $q = 0$ 的非空内部, 这里 $\mathcal{D}(C) = \{q \in \mathbb{R}^k \colon C(q) < \infty\}$.
3. $C(q)$ 为 $\mathbb{R}^k$ 上闭的凸函数.

B.3.9 注 (Ellis, 1984)

$C(q)$ 的定义域 $\mathcal{D}(C)$ 非空, 因 $C(0) = 0 < \infty$. 上述假设要求 $\mathcal{D}(C)$ 的内部非空. $C(q)$ 的凸性来自于 $C_n(q)$ 的凸性. 对于闭的要求, 即对于每个实数 b, $\{q \in \mathbb{R}^k \colon C(q) \leqslant b\}$ 在 $\mathbb{R}^k$ 中闭, 等价于 $C(q)$ 为下半连续. 若 $C(q)$ 是闭的, 并且 $\mathcal{D}(C)$ 为开集, 则 $C(q)$ 是陡峭的 (定义 B.2.1).

B.3.10 Gärtner-Ellis 定理 (Ellis, 1984, 定理 II.2)

令 Q_n 为 $Y_n = U_n/a_n$ 在 $\mathbb{R}^k$ 上的分布. 同样定义函数 $I(y)$ 为 $C(q)$ 的 Legendre-Fenchel 变换, 即

$$I(y)=\sup_{q\in\mathbb{R}^k}\{\langle q,y\rangle-C(q)\},\quad y\in\mathbb{R}^k. \tag{B.6}$$

若假设 B.3.8 成立, 则下列结论成立.

1. $I(y)$ 凸、闭 (上半连续)、非负. $I(y)$ 具有紧的水平集, 并且 $\inf_{y\in\mathbb{R}^k} I(y)=0$.
2. 偏差的上界, 即式 (B.3) 成立.
3. 若进一步假设 $C(q)$ 在其定义域的内部可微, 并且 $C(q)$ 是陡峭的 (定义 B.2.1), 则大偏差的下界, 即式 (B.4) 成立.

B.3.11 推论 (Ellis, 1985, 定理 II.4.1)

令 $X_1, X_2, \cdots$ 为一列取值在 $\mathbb{R}^k$ 中的独立同分布的随机变量. 令 $U_n=\sum_{i=1}^n X_i$, 并且 $C(q)$ 由式 (B.5) 给出. 假设对于所有 $q\in\mathbb{R}^k$, $C(q)$ 有限, 并如式 (B.6) 定义 $I(y)$. 则下列结论成立.

1. $Y_n=U_n/n$ 在 $\mathbb{R}^k$ 上的分布 $\{Q_n\}$, 对于 $a_n=n$ 及熵函数 $I(y)$ 具有大偏差性质.
2. $I(y)$ 为关于 y 的凸函数. 它量度 y 和 $E[X_1]$ 之间的差别, 即 $I(y)\geqslant 0$, 等号成立当且仅当 $y=E[X_1]$.
3. 若 X_1 的定义域为有限集 $\{x_1,x_2,\cdots,x_r\}\subseteq\mathbb{R}$, 其中 $x_1<x_2<\cdots<x_r$, 则当 $y\in[x_1,x_r]$, $I(y)$ 有限且连续, 并且当 $y\notin[x_1,x_r]$, $I(y)=\infty$.

B.3.12 引理 (Ellis, 1985, 214 页)

在 $\mathbb{R}$ $(k=1)$ 上的凸函数 f 为闭的当且仅当它在 $\mathcal{D}(f)$ 上连续, 并且

1. 在 $\mathcal{D}(f)$ 的端点上连续, 当端点属于 $\mathcal{D}(f)$, 即 $\mathcal{D}(f)$ 为闭集, 或
2. $f(x)\to\infty$ 当 x 趋于任何不在 $\mathcal{D}(f)$ 中的端点, 即 $\mathcal{D}(f)$ 为开集, 并且 f 是陡峭的.

B.3.13 定义——指数速度收敛 (Ellis, 1985, 231 页)

令 y_0 为 $\mathbb{R}^k$ 中的点. 称 $Y_n=U_n/a_n$ 指数速度收敛到 y_0, 记为

$$Y_n\xrightarrow{\exp} y_0,$$

若对于任意 $\varepsilon>0$, 存在 $N=N(\varepsilon)>0$, 使得对于充分大的 n 有

$$\rho_n\{\omega\colon \|Y_n(\omega)-y_0\|\geqslant\varepsilon\}\leqslant\exp(-a_nN).$$

B.3.14 定理 (Ellis, 1984, 定理 IV.1)

若假设 B.3.8 成立, 则下面各条等价.

1. $Y_n = U_n/a_n \xrightarrow{\text{exp}} y_0$.
2. $C(q)$ 在 $q=0$ 可微, 并且 $\nabla C(0) = y_0$.
3. 在 $\mathbb{R}^k$ 上, $I(y)$ 只在点 $y = y_0$ 取到的最小值.

B.3.15 定理 (Ellis, 1985, 49 页)

设随机向量 $\{U_n; n = 1, 2, \cdots\}$ 都定义在同一概率空间 $(\Omega, \mathcal{F}, \rho)$ 上. 若对所有 $N > 0$, $\sum_{n=1}^{\infty} \exp(-a_n N) < \infty$, 则 $Y_n = U_n/a_n \xrightarrow{\text{exp}} y_0$ 可推出 $Y_n = U_n/a_n \xrightarrow{\text{a.s.}} y_0$.

B.3.16 例

令 $\{X_i\}$, $i = 1, 2, \cdots$ 为独立正规随机变量, 具有平均值 μ 及单位方差. 定义 U_n 为关于 n 的部分和

$$U_n = X_1 + \cdots + X_n.$$

令重比例化常数为 $a_n = n$, 并且 $Y_n = U_n/n$, 则 U_n 的重比例化累积生成函数为

$$\begin{aligned} C_n(q) &= \frac{1}{a_n} \log E_n[e^{qU_n}] \\ &= \frac{1}{n} \log E[e^{qX_1}]^n \\ &= \frac{q^2}{2} + \mu q, \quad q \in \mathbb{R}. \end{aligned}$$

所以

$$C(q) = \lim_{n\to\infty} C_n(q) = \frac{q^2}{2} + \mu q.$$

现在的问题等价于例 B.2.4. 由定理 B.3.14 可得

$$Y_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{\text{exp}} C'(0) = \mu, \quad n \to \infty.$$

从而得到了弱大数定律, 此外还证明收敛具有指数速度. 因为

$$\sum_{n=1}^{\infty} e^{-nN} = \frac{e^{-N}}{1 - e^{-N}} < \infty, \quad \forall N > 0,$$

从而由定理 B.3.15 可得强大数定律, 即

$$Y_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{\text{a.s.}} \mu, \quad n \to \infty.$$

B.3.17　定理 (Ellis, 1984, 定理 V.1)

若假设 B.3.8 成立, 则下列结论成立.

1. $I(y)$ 为 $\mathbb{R}^k$ 上闭的凸函数.

2. $\langle q, y\rangle \leqslant C(q) + I(y)$, 对于所有 $q \in \mathcal{D}(C)$ 及所有 $y \in \mathcal{D}(I)$.

3. $\langle q, y\rangle = C(q) + I(y)$ 当且仅当 $y \in \partial C(q)$.

4. $y \in \partial C(q)$ 当且仅当 $q \in \partial I(y)$.

5. $C(q) = \sup_{y\in\mathbb{R}^k}\{\langle q, y\rangle - I(y)\}$, 对于所有 $q \in \mathbb{R}^k$.

6. 对于每个实数 b, $\{y\colon I(y) \leqslant b\}$ 为 $\mathbb{R}^k$ 的一个闭的、有界的、凸的子集.

7. $\inf_{y\in\mathbb{R}^k} I(y) = 0$, 并且 $I(y_0) = 0$ 当且仅当 $y_0 \in \partial C(q)$, 这是 $\mathbb{R}^k$ 的一个非空的、闭的、有界的、凸子集.

参考文献

Abarbanel, H.D.I. (1995). *Analysis of Observed Chaotic Data.* Springer-Verlag, New York.

Abarbanel, H.D.I., Brown, R., Sidorowich, J.J. and Tsimring, L.S. (1993). The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.* **65**(4), 1331–1392.

Abramowitz, M. and Stegun, I.A. (1964). *Handbook of Mathematical Functions.* Dover, New York.

Adler, R.J. (1981). *The Geometry of Random Fields.* John Wiley & Sons, Chichester.

Aki, K. (1989). Ideal probabilistic earthquake prediction. *Tectonophysics* **169**, 197–198.

Arbeiter, M. (1991). Random recursive construction of self-similar fractal measures. The compact case. *Probab. Th. Rel. Fields* **88**, 497–520.

Bebbington, M. and Harte, D.S. (2001). On the statistics of the linked stress release model. *J. Appl. Prob.* **38A**, 186–198.

Beck, C. (1990). Upper and lower bounds on the Rényi dimensions and the uniformity of multifractals. *Physica D* **41**, 67–68.

Beran, J. (1994). *Statistics for Long Memory Processes.* Chapman & Hall, New York.

Berliner, L.M. (1992). Statistics, probability and chaos. *Statistical Sci.* **7**(1), 69–122.

Berman, S.M. (1970). Gauss processes with stationary increments: local times and sample function properties. *Ann. Statist.* **41**(4), 1260–1272.

Billingsley, R (1965). *Ergodic Theory and Information.* John Wiley & Sons, New York.

Bohr, T., Jensen, M.H., Paladin, G. and Vulpiani, A. (1998). *Dynamical Systems Approach to Turbulence. Cambridge Nonlinear Science Series.* Cambridge University Press, Cambridge.

Bouchaud, J-R, Potters, M. and Meyer, M. (2000). Apparent multifractality in financial time series. *European Physical J. B* **13**, 595–599.

Breiman, L. (1968). *Probability.* Addison Wesley, Reading, MA.

Brown, G., Michon, G. and Peyrière, J. (1992). On the multifractal analysis of measures. *J. Statist. Physics* **66**(3/4), 775–790.

Bucklew, J.A. (1990). *Large Deviation Techniques in Decision, Simulation and Estimation.* John Wiley & Sons, New York.

Casdagli, M., Eubank, S., Farmer, J.D. and Gibson, J. (1991). State space reconstruction in the presence of noise. *Physica D* **51**(1-3), 52–98.

Cawley, R. and Mauldin, R.D. (1992). Multifractal decompositions of Moran fractals. *Adv. Math.* **92**, 196–236.

Chan, G. and Wood, A.T.A. (2000). Increment-based estimators of fractal dimension for two-dimensional surface data. *Statistica Sinica* **10**, 343–376.

Chatterjee, S. and Yilmaz, M.R. (1992). Chaos, fractals and statistics (with discussion). *Statistical Sci.* **7**(1), 49–68.

Cheng, B. and Tong, H. (1994). Orthogonal projection, embedding dimension and sample size in chaotic time series from a statistical perspective. *Phil. Trans. Royal Soc. Lond. A* **348**(1688), 325–341.

Ciesielski, E. and Taylor, S.J. (1962). First passage times and sojourn times for Brown motion in space and the exact Hausdorff measure of the sample path. *Trans. Amer. Math. Soc.* **103**, 434–450.

Constantine, A.G. and Hall, P. (1994). Characterizing surface smoothness via estimation of effective fractal dimension. *J. R. Statist. Soc. B* **56**(1), 97–113.

Cutler, CD. (1986). The Hausdorff dimension distribution of finite measures in Euclidean space. *Canad. J. Math.* **38**(6), 1459–1484.

Cutler, CD. (1991). Some results on the behaviour and estimation of the fractal dimensions of distributions on attractors. *J. Statist. Physics* **62**(3/4), 651–708.

Cutler, CD. (1994). A theory of correlation dimension for stationary time series. *Phil. Trans. Royal Soc. Lond. A* **348**(1688), 343–355.

Cutler, CD. (1995). Strong and weak duality principles for fractal dimension in Euclidean space. *Math. Proc. Camb. Phil. Soc.* **118**(3), 393–410.

Cutler, CD. (1997). A general approach to predictive and fractal scaling dimensions in discrete-index time series. *Fields Instit. Comm.* **11**, 29–48.

Cutler, CD. and Dawson, D.A. (1989). Estimation of dimension for spatially distributed data and related limit theorems. *J. Multivar. Anal.* **28**, 115–148.

Cutler, CD. and Kaplan, D.T. (1997). *Nonlinear Dynamics and Time Series: Building a Bridge Between the Natural and Statistical Sciences. Fields Institute Communications Volume 11.* American Mathematical Society, Providence, RI.

Cvitanović, P. (1993). *Universality in Chaos. 2nd Edition.* Institute of Physics Publishing, Bristol.

Daley, D.J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* Springer-Verlag, New York.

David, H.A. (1970). *Order Statistics.* John Wiley & Sons, New York.

Davies, R.B. (1980). The distribution of a linear combination of χ^2 random variables. Algorithm AS 155. *Appl. Statist.* **29**(3), 323–333.

Davies, R.B. and Harte, D.S. (1987). Tests for Hurst effect. *Biometrika* **74**(1), 95–101.

Davies, S. and Hall, P. (1999). Fractal analysis of surface roughness by using spatial data (with discussion). *J. R. Statist. Soc. B* **61**(1), 3–37.

Denker, M. and Keller, G. (1986). Rigorous statistical procedures for data from dynamical systems. *J. Statist. Physics* **44**(1/2), 67–93.

Dobrushin, R.L. (1979). Gauss and their subordinated self-similar random generalized fields. *Ann. Prob.* **7**(1), 1–28.

Dobrushin, R.L. and Major, P. (1979). Noncentral limit theorems for nonlinear functionals of Gauss fields. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **50**, 27–52.

Dvorak, I. and Klaschka, J. (1990). Modification of the Grassberger-Procaccia algorithm for estimating the correlation exponent of chaotic systems with high embedding dimension. *Physics Letters A* **145**(5), 225–231.

Eckmann, J.-P. and Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**(3), 617–656.

Eckmann, J.-P. and Ruelle, D. (1992). Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems. *Physica D* 56(2/3), 185–187.

Edgar, G.A. and Mauldin, R.D. (1992). Multifractal decompositions of digraph recursive fractals. *Proc. London Math. Soc.* **65**, 604–628.

Efron, B. and Tibshirani, R.J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Sci.* **1**(1), 54–77.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, New York.

Eggleston, H.G. (1949). The fractional dimension of a set defined by decimal properties. *Quart. J. Math. (Oxford)* **20**, 31–36.

Ellis, R.S. (1984). Large deviations for a general class of random vectors. *Ann. Prob.* **12**(1), 1–12.

Ellis, R.S. (1985). *Entropy, Large Deviations, and Statistical Mechanics.* Springer-Verlag, New York.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, New York.

Eneva, M. (1996). Effect of limited data sets in evaluating the scaling properties of spatially distributed data: an example from mining induced seismic activity. *Geophys. J. Int.* **124**(3), 773–786.

Essex, C. and Nerenberg, M.A.H. (1991). Comments on 'Deterministic chaos: the science and the fiction' by D. Ruelle. *Proc. R. Soc. Lond. A* **435**(1894), 287–292.

Falconer, K.J. (1985). *The Geometry of Fractal Sets.* Cambridge University Press, Cambridge.

Falconer, K.J. (1986). Random fractals. *Math. Proc. Comb. Phil. Soc.* **100**(3), 559–582.

Falconer, K.J. (1990). *Fractal Geometry, Mathematical Foundations and Applications.* John Wiley & Sons, Chichester.

Falconer, K.J. (1994). The multifractal spectrum of statistically self-similar measures. *J. Theor. Probab.* **7**(3), 681–702.

Falconer, K.J. (1997). *Techniques in Fractal Geometry.* John Wiley & Sons, Chichester.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications. Volume 2, 2nd Edition.* John Wiley & Sons, New York.

Feuerverger, A., Hall, P. and Wood, A.T.A. (1994). Estimation of fractal index and fractal dimension of a Gauss process by counting the number of level crossings. *J. Time Series Anal.* **15**(6), 587–606.

Fisher, A., Calvet, L. and Mandelbrot, B.B. (1997a). Large deviations and the distribution of prices changes. Cowles Foundation Discussion Paper 1165. Yale University, New Haven.

Fisher, A., Calvet, L. and Mandelbrot, B.B. (1997b). Multifractality of Deutschemark / US Dollar exchange rates. Cowles Foundation Discussion Paper 1166. Yale University, New Haven.

Fraser, A.M. and Swinney, H.L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Rev. A* **33**(2), 1134–1140.

Friedlander, S.K. and Topper, L. (1961). *Turbulence: Classic Papers on Statistical Theory.* Interscience Publishers, New York.

Frisch, U. (1991). From global scaling, à la Kolmogorov, to multifractals scaling in fully developed turbulence. *Proc. R. Soc. Lond. A* **434**(1890), 89–99.

Fristedt, B. (1974). Sample functions of stochastic processes with stationary independent increments. In: *Advances in Probability. Vol 3*, 241–393. (Edited by P. Ney and S.C. Port). Marcel Dekker, New York.

Gács, P. (1973). Hausdorff-dimension and probability distributions. *Period. Math. Hungarica* **3**(1/2), 59–71.

Gärtner, J. (1977). On large deviations from the invariant measure. *Theor. Prob. Applic.* **22**(1), 24–39.

Geilikman, M.B., Golubeva, T.V. and Pisarenko, V.F. (1990). Multifractal patterns of seismicity. *Earth & Plan. Sci. Let.* **99**, 127–132.

Geller, R.J. (1997). Earthquake prediction: a critical review. *Geophys. J. Int.* **131**(3), 425–450.

Geller, R.J., Jackson, D.D., Kagan, Y.Y. and Mulargia, F. (1997). Earthquakes cannot be predicted. *Science* **275**(5306), 1616–1617.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Series Anal.* **4**(4), 221–238.

Graf, S. (1987). Statistically self-similar fractals. *Probab. Th. Rel. Fields* **74**(3), 357–392.

Grassberger, P. (1983). Generalized dimensions of strange attractors. *Physics Letters* **97A**(6), 227–230.

Grassberger, P. and Procaccia, I. (1983a). Measuring the strangeness of strange attractors. *Physica D* **9**(1/2), 189–208.

Grassberger, P. and Procaccia, I. (1983b). Estimation of the Kolmogorov entropy from a chaotic signal. *Physical Rev. A* **28**(4), 2591–2593.

Grassberger, P. and Procaccia, I. (1983c). Characterisation of strange attractors. *Physical Rev. Utters* **50**(5), 346–349.

Grebogi, C, Ott, E. and Yorke, J.A. (1988). Roundoff-induced periodicity and the correlation dimension of chaotic attractors. *Physical Rev. A* **38**(7), 3688–3692.

Gupta, Y.K. and Waymire, E.C. (1990). Multiscaling properties of spatial rainfall and river flow distributions. *J. Geophys. Res.* **95**(D3), 1999–2009.

Gupta, Y.K. and Waymire, E.C. (1993). A statistical analysis of mesoscale rainfall as a random cascade. *J. Appl. Meteorology* **32**(2), 251–267.

Halsey, T.C., Jensen, M.H., Kadanoff, L.P., Procaccia, I. and Shraiman, B.I. (1986). Fractal measures and their singularities: the characterisation of strange sets. *Physical Rev. A* **33**(2), 1141–1151.

Hao, B.L. (1990). *Chaos II.* World Scientific Publishing Company, Singapore.

Harte, D.S. (1998). Dimension estimates of earthquake epicentres and hypocentres. *J. Nonlinear Sci.* **8**, 581–618.

Harte, D.S. and Vere-Jones, D. (1999). Differences in coverage between the PDE and New Zealand local earthquake catalogues. *NZ. J. Geology & Geophysics* **42**, 237–253.

Haslett, J. and Raftery, A.E. (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind power resource. *Appl. Statist.* **38**(1), 1–50.

Hentschel, H.G.E. and Procaccia, I. (1983). The infinite number of generalized dimensions of fractals and strange attractors. *Physica D* **8**(3), 435–444.

Heyde, C.C. (1999). A risky asset model with strong dependence through fractal activity time. *J. Appl. Prob.* **36**, 1234–1239.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**(5), 1163–1174.

Hirata, T. and Imoto, M. (1991). Multifractal analysis of spatial distribution of micro-earthquakes in the Kanto region. *Geophys. J. Int.* **107**(1), 155–162.

Holley, R. and Waymire, E.C. (1992). Multifractal dimensions and scaling exponents for strongly bounded random cascades. *Ann. Appl. Prob.* **2**(4), 819–845.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika* **68**(1), 165–176.

Hunt, J.C.R., Phillips, O.M. and Williams, D. (1991). Turbulence and stochastic processes: Kolmogorov's ideas 50 years on. Preface and contents to this special issue. *Proc. R. Soc. Lond. A* **434**(1890), 5–7.

Hurst, H.E. (1951). Long term storage capacity of reservoirs. *Trans. Amer. Soc. of Civil Engineers* **116**, 770–808.

Hutchinson, J.E. (1981). Fractals and self similarity. *Indiana Uni. Math. J.* **30**(5), 713–747.

Isham, V. (1993). Statistical aspects of chaos: a review. In: *Networks and Chaos-Statistical and Probabilistic Aspects*, 124–200. (Edited by O. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall). Chapman & Hall, London.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions. Volume 2. Second Edition.* John Wiley & Sons, New York.

Kagan, Y.Y. (1981a). Spatial distribution of earthquakes: the three point moment function. *Geophys. J. R. Astron. Soc.* **67**, 697–717.

Kagan, Y.Y. (1981b). Spatial distribution of earthquakes: the four point moment function. *Geophys. J. R. Astron. Soc.* **67**, 719–733.

Kagan, Y.Y. (1991). Fractal dimension of brittle fracture. *J. Nonlinear Sci.* **1**, 1–16.

Kagan, Y.Y. (1997). Are earthquakes predictable? *Geophys. J. Int.* **131**(3), 505–525.

Kagan, Y.Y. (1999). Is earthquake seismology a hard, quantitative science? *PAGEOPH* **155**, 233–258.

Kagan, Y.Y. and Knopoff, L. (1980). Spatial distribution of earthquakes: the two point correlation function. *Geophys. J. R. Astron. Soc.* **62**, 303–320.

Kahane, J.-P. and Peyrière, J. (1976). Sur certaines martingales de Benoit Mandelbrot. *Adv. Math.* **22**(2), 131–145.

Kennel, M.B. and Isabelle, S. (1992). Method to distinguish possible chaos from colored noise and to determine embedding parameters. *Physical Rev. A* **46**(6), 3111–3118.

Kolmogorov, A.N. (1941). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Comptes Rendus (Doklady) de I'Aacdemie des Sciences de I'U.R.S.S.* **30**, 301–305. Reprinted in: Friedlander & Topper (1961).

Kolmogorov, A.N. (1962). A refinement of previous hypotheses concerning the local struc- ture of turbulence in a viscous incompressible fluid at high Reynolds number. *J. Fluid Mechanics* **13**(1), 82–85.

Kolmogorov, A.N. and Gnedenko, B.V. (1954). *Limit Distributions for Sums of Independent Random Variables. (Translated by K.L. Chung.)* Addison Wesley, Reading, MA.

Lamperti, J. (1962). Semi-stable stochastic processes. *Trans. Amer. Math. Soc.* **104**, 62–78.

Lawrance, A.J. and Kottegoda, N.T. (1977). Stochastic modelling of river flow time series. *J. R. Statist. Soc. A* **140**(1), 1–31.

Lay, T. and Wallace, T.C. (1995). *Modern Global Seismology.* Academic Press, San Diego.

Li, W.K. and McLeod, A.I. (1986). Fractional time series modelling. *Biometrika* **73**(1), 217–221.

Lloyd, E.H. (1981). Stochastic hydrology: An introduction to wet statistics for dry statisticians. *Comm. Statist. A* **10**(15), 1505–1522.

Lorenz, E.N. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141.

Lorenz, E.N. (1993). *The Essence of Chaos.* UCL Press, London.

Lovejoy, S. and Schertzer, D. (1985). Generalized scale invariance in the atmosphere and fractal models of rain. *Water Resources Res.* **21**(8), 1233–1250.

Lovejoy, S. and Schertzer, D. (1990). Multifractals, universality classes and satellite and radar measurements of cloud and rain fields. *J. Geophys. Res.* **95**(D3),

2021–2034.

Lu, C.S., Harte, D.S. and Bebbington, M. (1999). A linked stress release model for historical Japanese earthquakes: coupling among major seismic regions. *Earth Planets Space* **51**,907–916.

Lukacs, E. (1960). *Characteristic Functions.* Griffin, London.

Major, P. (1981). Limit theorems for nonlinear functionals of Gauss sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **57**, 129–158.

Mandelbrot, B.B. (1974). Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *J. Fluid Mechanics* **62**(2), 331–358.

Mandelbrot, B.B. (1975). Limit theorems on the self normalised range for weakly and strongly dependent processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **31**, 271–285.

Mandelbrot, B.B. (1977). *Fractals: Form, Chance and Dimension.* Freeman, San Francisco.

Mandelbrot, B.B. (1983). *The Fractal Geometry of Nature.* Freeman, New York.

Mandelbrot, B.B. (1989). Multifractal measures, especially for the geophysicist. *PAGEOPH* **131**(1/2), 5–42.

Mandelbrot, B.B. (1990a). Negative fractal dimensions and multifractals. *Physica A* **163**, 306–315.

Mandelbrot, B.B. (1990b). New “anomalous” multiplicative multifractals: left sided f(a) and the modelling of DLA. *Physica A* **168**, 95–111.

Mandelbrot, B.B. (1991). Random multifractals: negative dimensions and the resulting limitations of the thermodynamic formalism. *Proc. R. Soc. Lond. A* **434**(1890), 79–88.

Mandelbrot, B.B. (1997). *Fractals and Scaling in Finance : Discontinuity, Concentration, Risk : Selecta Volume E.* Springer-Verlag, New York.

Mandelbrot, B.B. (1998). *Multifractals and $1/f$ Noise: Wild Self-Affinity in Physics (1963–1976): Selecta Volume N.* Springer-Verlag, New York.

Mandelbrot, B.B. (1999). Renormalization and fixed points in finance, since 1962. *Physica A* **263**(1), 477–487.

Mandelbrot, B.B., Evertsz, C.J.G. and Hayakawa, Y (1990). Exactly self-similar left-sided multifractal measures. *Physical Rev. A* **42**(8), 4528–4536.

Mandelbrot, B.B., Fisher, A. and Calvet, L. (1997). A multifractal model of asset returns. Cowles Foundation Discussion Paper 1164. Yale University, New Haven.

Mandelbrot, B.B. and Riedi, R.H. (1997). Inverse measures, the inversion formula, and discontinuous multifractals. *Adv. Appl. Math.* **18**(1), 50–58.

Mandelbrot, B.B. and Van Ness, J.W. (1968). Fractional Brown motions, fractional noises and applications. *SIAM Review* **10**(4), 422–437.

Mandelbrot, B.B. and Wallis, J.R. (1969). Robustness of the rescaled range R/S in the measurement of non-cyclic long run statistical dependence. *Water Resources Res.* **5**(5), 967–988.

Marcus, M.B. (1976). Capacity of level sets of certain stochastic processes. *Z Wahrscheinlichkeitstheorie verw. Gebiete* **34**, 279–284.

Mason, D.M. (1982). Laws of large numbers for sums of extreme values. *Ann. Prob.* **10**(3), 754–764.

Mauldin, R.D. and Williams, S.C. (1988). Random recursive constructions: asymptotic geometric and topological properties. *Trans. Amer. Math. Soc.* **295**(1), 325–346.

Maunder, D.E. (1994). New Zealand Seismological Report 1992. Seismological Observatory Bulletin E-176. Institute of Geological and Nuclear Sciences Report 94/47. Institute of Geological and Nuclear Sciences, Lower Hutt.

May, R.M. (1976). Simple mathematical models with very complicated dynamics. *Nature* **261**, 459–467.

May, R.M. (1987). Chaos and the dynamics of biological populations. *Proc. R. Soc. Lond. A* **413**(1844), 27–44.

Meneveau, C. and Sreenivasan, K.R. (1991). The multifractal nature of turbulent energy dissipation. *J. Fluid Mechanics* **224**, 429–484.

Mikosch, T. and Wang, Q. (1993). Some results on estimating Rényi type dimensions. Institute of Statistics and Operations Research Report. Victoria University of Wellington, Wellington.

Mikosch, T. and Wang, Q. (1995). A Monte-Carlo method for estimating the correlation exponent. *J. Statist. Physics* **78**(3/4), 799–813.

Molchan, G.M. (1995). Multifractal analysis of Brown zero set. *J. Statist. Physics* **79**(3/4), 701–730.

Monin, A.S. and Yaglom, A.M. (1971). *Statistical Fluid Mechanics; Mechanics of Turbulence.* MIT, Cambridge, MA.

Moran, RA.P (1946). Additive functions of intervals and Hausdorff measure. *Proc. Camb. Phil. Soc.* **42**, 15–23.

Moran, RA.R (1950). Some remarks on animal population dynamics. *Biometrics* **6**(3), 250–258.

Morandi, M.T. and Matsumura, S. (1991). Update on the examination of the seismic observational network of the National Research Institute for Earth Science and Disaster Prevention (NIED): detection capability and magnitude correction. No. 47, 1–18. National Research Institute for Earth Science and Disaster Prevention, Tsukuba.

Nerenberg, M.A.H. and Essex, C. (1990). Correlation dimension and systematic geometric effects. *Physical Rev. A* **42**(12), 7065–7074.

Nicolis, C. and Nicolis, G. (1984). Is there a climate attractor? *Nature* **311**, 529–532.

Oboukhov, A.M. (1962). Some specific features of atmospheric turbulence. *J. Fluid Mechanics* **13**(1), 77–81.

Ogata, Y. and Abe, K. (1991). Some statistical features of the long term variation of the global and regional seismic activity. *Int. Statist. Rev.* **59**(2), 139–161.

Olsen, L. (1994). *Random Geometrically Graph Directed Self-Similar Multifractals.* Longman Scientific & Technical, Harlow, Essex.

Orey, S. (1970). Gauss sample functions and the Hausdorff dimension of level crossings. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **15**, 249–256.

Osborne, A.R. and Provenzale, A. (1989). Finite correlation dimension for stochastic systems with power-law spectra. *Physica D* **35**(3), 357–381.

Ott, E., Sauer, T. and Yorke, J.A. (1994). *Coping with Chaos. Analysis of Chaotic Data with the Exploitation of Chaotic Systems.* John Wiley & Sons, New York.

Pachard, N.H., Crutchfield, J.P., Farmer, J.D. and Shaw, R.S. (1980). Geometry from atime series. *Physical Rev. Letters* **45**(9), 712–716.

Paladin, G. and Vulpiani, A. (1987). Anomalous scaling laws in multifractal objects. *Physics Reports* **156**(4), 147–225.

Papanastassiou, D. and Matsumura, S. (1987). Examination of the NRCDP's (The National Research Center for Disaster Prevention) seismic observational network as regards: I. detectability-locatability, II. accuracy of the determination of the earthquake source parameters. No. 39, 37–65. National Research Institute for Earth Science and Disaster Prevention, Tsukuba.

Park, K. and Willinger, W. (Editors) (2000). *Self-Similar Network Traffic and Performance Evaluation.* John Wiley & Sons, New York.

Pesin, Y.B. (1993). On rigorous mathematical definitions of correlation dimension and generalized spectrum for dimensions. *J. Statist. Physics* **71**(3/4), 529–547.

Pisarenko, D.V. and Pisarenko, V.F. (1995). Statistical estimation of the correlation dimension. *Physics Letters A* **197**(1), 31–39.

Porter-Hudak, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *J. Amer. Statist. Assoc.* **85**(410), 338–344.

Rasband, S.N. (1990). *Chaotic Dynamics of Nonlinear Systems.* John Wiley & Sons, New York.

Rényi, A. (1959). On the dimension and entropy of probability distributions. *Ada Mathematica* **10**, 193–215.

Rényi, A. (1965). On the foundations of information theory. *Review of I.S.I.* **33**(1), 1–14.

Rényi, A. (1970). *Probability Theory.* Elsevier Science Publishers B.V. (North-Holland), Amsterdam.

Resnick, S.I. (1997). Heavy tail modeling and teletraffic data. *Ann. Statist.* **25**(5), 1805–1869.

Riedi, R. (1995). An improved multifractal formalism and self-similar measures. *J. Math. Anal. Applic.* **189**(2), 462–490.

Riedi, R.H. (2001). Multifractal processes. In: *Long-Range Dependence: Theory and Applications.* (Edited by: by Murad S. Taqqu, George Oppenheim, Paul Doukhan). Birkhauser, Boston.

Riedi, R.H. and Mandelbrot, B.B. (1995). Multifractal formalism for infinite multinomial measures. *Adv. Appl. Math.* **16**(2), 132–150.

Riedi, R.H. and Mandelbrot, B.B. (1997). Inversion formula for continuous multifractals. *Adv. Appl. Math.* **19**(3), 332–354.

Riedi, R.H. and Mandelbrot, B.B. (1998). Exceptions to the multifractal formalism for discontinuous measures. *Math. Proc. Camb. Phil. Soc.* **123**(1), 133–157.

Riedi, R.H. and Willinger, W. (2000). Toward an improved understanding of network traffic dynamics. In: *Self-Similar Network Traffic and Performance Evaluation.* (Edited by: Kihong Park and Walter Willinger). John Wiley & Sons, New York.

Rogers, C.A. (1970). *Hausdorff Measures.* Cambridge University Press, Cambridge.

Rosenblatt, M. (1961). Independence and dependence. In: *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability. Volume II. Probability Theory*, 431–443. (Edited by J. Neyman). University of California Press, Berkley.

Rosenblatt, M. (1979). Some limit theorems for partial sums of quadratic forms in station- ary Gauss variables. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **49**,

125–132.

Rosenblatt, M. (1981). Limit theorems for Fourier transforms of functionals of Gauss sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **55**, 123–132.

Ruelle, D. (1989). *Chaotic Evolution and Strange Attractors.* Cambridge University Press, Cambridge.

Ruelle, D. (1990). The Claude Bernard Lecture, 1989. Deterministic chaos: the science and the fiction. *Proc. R. Soc. Lond. A* **427**(1873), 241–248.

Ruelle, D. (1995). *Turbulence, Strange Attractors, and Chaos.* World Scientific Publishing Company, Singapore.

Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gauss Random Processes: Stochastic Models with Infinite Variance.* Chapman & Hall, New York.

Sauer, T., Yorke, J.A. and Casdagli, M. (1991). Embedology. *J. Statist. Physics* **65**(3/4), 579–616.

Schertzer, D. and Lovejoy, S. (1987). Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *J. Geophys. Res.* **92**(D8), 9693–9714.

Schertzer, D. and Lovejoy, S. (1989). Generalised scale invariance and multiplicative processes in the atmosphere. *PAGEOPH* **130**(1), 57–81.

Scholz, C.H. (1990). *The Mechanics of Earthquakes and Faulting.* Cambridge University Press, Cambridge.

Scholz, C.H. and Mandelbrot, B.B. (1989). Fractals in Geophysics. *Reprint from PAGEOPH* **131**(1/2). Birkhauser Verlag, Basel.

Smith, R.L. (1992a). Optimal estimation of fractal dimension. In: *Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity. Proceedings of the Workshop on Nonlinear Modeling and Forecasting, September 1990, at Santa Fe, New Mexico. Vol XII,* 115–135. (Edited by M. Casdagli and S. Eubank). Addison Wesley, Redwood City, CA.

Smith, R.L. (1992b). Estimating dimension in noisy chaotic time series. *J. R. Statist. Soc. B* **54**(2), 329–351.

Sphilhaus, A. (1991). *Atlas of the World and Geophysical Boundaries Showing Oceans, Continents and Tectonic Plates in Their Entirety.* American Philosophical Society, Philadelphia.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics.* John Wiley & Sons, Chichester.

Takens, F. (1981). Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence, Warwick 1980. Proceedings of a Symposium held at the University of Warwick 1979/80. Lecture Notes in Mathematics 898*, 366–381. (Edited by D.A. Rand and L.-S. Young). Springer-Verlag, Berlin.

Takens, F. (1985). On the numerical determination on the dimension of an attractor. In: *Dynamical Systems and Bifurcations. Proceedings of a workshop held in Groningen, The Netherlands, April 16–20, 1984. Lecture Notes in Mathematics, 1125*, 99–106. (Edited by B.L.J. Braaksma, H.W. Broer and F. Takens). Springer-Verlag, Berlin.

Takens, F. (1993). Detecting nonlinearities in stationary time series. *Int. J. Bifurcation and Chaos* **3**(2), 241–256.

Taqqu, M.S. (1975). Weak convergence to fractional Brown motion and to the Rosenblatt process. *Z Wahrscheinlichkeitstheorie verw. Gebiete* **31**, 287–302.

Taqqu, M.S. (1977). Law of the iterated logarithm for sums of nonlinear functions of Gauss variables that exhibit long range dependence. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **40**, 203–238.

Taqqu, M.S. (1978). A representation for self-similar processes. *Stochastic Processes & Appl.* **7**, 55–64.

Taqqu, M.S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Z Wahrscheinlichkeitstheorie verw. Gebiete* **50**, 53–83.

Taqqu, M.S., Teverovsky, V and Willinger, W. (1995). Estimators for long-range dependence: An empirical study. *Fractals* **3**(4), 785–798.

Taqqu, M.S. and Wolpert, R.L. (1983). Infinite variance self-similar processes subordinate to a Poisson measure. *Z. Wahrscheinlichkehstheorie verw. Gebiete* **62**, 53–72.

Taylor, S.J. (1964). The exact Hausdorff measure of the sample path for planar Brown motion. *Proc. Camb. Phil. Soc.* **60**, 253–258.

Taylor, S.J. (1986). The measure theory of random fractals. *Math. Proc. Camb. Phil. Soc.* **100**(3), 383–406.

Theiler, J. (1986). Spurious dimension from correlation algorithms applied to limited time-series data. *Physical Rev. A* **34**(3), 2427–2432.

Theiler, J. (1988). Lacunarity in a best estimator of fractal dimension. *Physics Letters A* **133**(4/5), 195–200.

Theiler, J. (1990). Statistical precision of dimension estimators. *Physical Rev. A* **41**(6), 3038–3051.

Theiler, J. (1991). Some comments on the correlation dimension of l//a noise. *Physics letters A* **155**(8/9), 480–493.

Tricot, C. (1982). Two definitions of fractal dimension. *Math. Proc. Camb. Phil. Soc.* **91**, 57–74.

Tricot, C. (1995). *Curves and Fractal Dimension.* Springer-Verlag, New York.

Vere-Jones, D. (1995). Forecasting earthquakes and earthquake risk. *Int. J. Forecasting* **11**, 503–538.

Vere-Jones, D. (1999). On the fractal dimension of point patterns. *Adv. Appl. Prob.* **31**, 643–663.

Vere-Jones, D. (2000). Seismology - a statistical vignette. *J. Amer. Statist. Assoc.* **95**(451), 975–978.

Walters, P. (1982). *An Introduction to Ergodic Theory.* Springer-Verlag, New York.

Wang, Q. (1995). Fractal dimension estimates of global and local temperature data. *J. Appl. Meteorology* **34**(11), 2556–2564.

Wells, D. and Coppersmith, K.J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull. Seismol. Soc. Amer.* **84**(4), 974–1002.

Whitney, H. (1936). Differentiable manifolds. *Ann. Mathematics* **37**(3), 645–680.

Willinger, W. and Paxson, V (1998). Where mathematics meets the internet. *Notices of the American Mathematical Society* **45**(8), 961–970.

Willinger, W., Taqqu, M.S., Leland, W.E. and Wilson, D.V (1995). Self-similarity in high speed packet traffic: analysis and modeling of ethernet traffic measurements. *Statistical Sci.* **10**(1), 67–85.

Wolff, R.C.L. (1990). A note on the behaviour of the correlation integral in the presence of a time series. *Biometrika* **77**(4), 689–697.

Wyss, M., Aceves, R.L. and Park, S.K. (1997). Cannot earthquakes be predicted? (with response). *Science* **278**(5337), 487–490.

Young, L.-S. (1982). Dimension, entropy and Lyapunov exponents. *Ergod. Th. Dynam. Sys.* **2**, 109–124.

译 后 记

分形几何著名专家、清华大学文志英教授的中文版序可以作为本书的一个导读, 他不仅阐述了本书的主要内容及特点, 还对测度的重分形分析的来源、核心问题及与其他学科的联系等给了一个清晰的概述. 本书的确是一本将分形几何, 特别是将重分形分析理论应用于统计的非常好的参考书. 需要补充一点的是, 本书作者还为我们提供了大量有价值的参考文献.

本书由华南理工大学数学系分形几何及动力系统课题组共同翻译, 李兵译前言、记号列表及第 1 章; 吴敏译第 2 章; 匡锐译第 3、4 章; 李进军译第 5、6 章; 马东魁译第 7 章; 王晓天译第 8 章; 章雄鹰译第 9, 11 章; 熊瑛译第 10 章; 娄曼丽译附录. 吴敏审阅了全书, 匡锐、熊瑛和李兵编译了书中的图表并编排了正文, 李彦哲博士编排了部分正文. 由于译者水平所限, 译文中难免有疏漏和错误之处, 敬请读者批评指正.

本书得到国家自然科学基金 (11071082, 11001086, 11101159) 及中央高校基础研究基金 (2012ZZ0073) 的支持, 译者衷心感谢文志英教授的鼓励及撰写的中文版序.

吴 敏

2012 年 1 月于华南理工大学

《现代数学译丛》已出版书目

（按出版时间排序）

1 椭圆曲线及其在密码学中的应用——导引 2007.12 〔德〕Andreas Enge 著 吴 铤 董军武 王明强 译
2 金融数学引论——从风险管理到期权定价 2008.1 〔美〕Steven Roman 著 邓欣雨 译
3 现代非参数统计 2008.5 〔美〕Larry Wasserman 著 吴喜之 译
4 最优化问题的扰动分析 2008.6 〔法〕J. Frédéric Bonnans 〔美〕Alexander Shapiro 著 张立卫 译
5 统计学完全教程 2008.6 〔美〕Larry Wasserman 著 张 波 等译
6 应用偏微分方程 2008.7 〔英〕John Ockendon, Sam Howison, Andrew Lacey & Alexander Movchan 著 谭永基 程 晋 蔡志杰 译
7 有向图的理论、算法及其应用 2009.1 〔丹〕J. 邦詹森 〔英〕G. 古廷 著 姚兵 张忠辅 译
8 微分方程的对称与积分方法 2009.1〔加〕乔治 W. 布卢曼 斯蒂芬 C. 安科 著 闫振亚 译
9 动力系统入门教程及最新发展概述 2009.8 〔美〕Boris Hasselblatt & Anatole Katok 著 朱玉峻 郑宏文 张金莲 阎欣华 译 胡虎翼 校
10 调和分析基础教程 2009.10 〔德〕Anton Deitmar 著 丁勇 译
11 应用分支理论基础 2009.12 〔俄〕尤里·阿·库兹涅佐夫 著 金成桴 译
12 多尺度计算方法——均匀化及平均化 2010.6 Grigorios A. Pavliotis, Andrew M. Stuart 著 郑健龙 李友云 钱国平 译
13 最优可靠性设计：基础与应用 2011.3 〔美〕Way Kuo, V. Rajendra Prasad, Frank A.Tillman, Ching-Lai Hwang 著 郭进利 闫春宁 译 史定华 校
14 非线性最优化基础 2011.4 〔日〕Masao Fukushima 著 林贵华 译
15 图像处理与分析：变分，PDE，小波及随机方法 2011.6 Tony F. Chan, Jianhong (Jackie) Shen 著 陈文斌，程晋 译
16 马氏过程 2011.6 〔日〕福岛正俊 竹田雅好 著 何萍 译 应坚刚 校

17 合作博弈理论模型 2011.7 〔罗〕Rodica Branzei 〔德〕Dinko Dimitrov 〔荷〕Stef Tijs 著 刘小冬 刘九强 译

18 变分分析与广义微分 I：基础理论 2011.9 〔美〕Boris S. Mordukhovich 著 赵亚莉 王炳武 钱伟懿 译

19 随机微分方程导论应用(第 6 版) 2012.4 〔挪〕Bernt Øksendal 著 刘金山 吴付科 译

20 金融衍生产品的数学模型 2012.4 郭宇权(Yue-Kuen Kwok) 著 张寄洲 边保军 徐承龙 等 译

21 欧拉图与相关专题 2012.4 〔英〕Herbert Fleischner 著 孙志人 李 皓 刘桂真 刘振宏 束金龙 译 张 昭 黄晓晖 审校

22 重分形：理论及应用 2012.5 〔美〕戴维·哈特 著 华南理工分形课题组 译

(O-4732.0101)

科学出版中心　数理分社
电话：(010)64033664
E-mail: math-phy@mail.sciencep.com
网址：http://www.math-phy.cn

销售分类建议：高等数学

www.sciencep.com



定价：65.00 元